EXPLORING HETEROGENEITY OF STATED PREFERENCES THROUGH LATENT CLASS ANALYSIS

by
Mo Zhou

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

April, 2017

Abstract

Patient preferences have been increasingly incorporated into clinical and regulatory decision-making. It leads to a growing interest in advancing and applying methods to study preference heterogeneity. Latent class analysis (LCA) is an emerging technique used in stated-preference studies to segment people by preferences instead of observed characteristics (e.g. demographics). The objective of this dissertation is to examine and advance the application of LCA in stated-preference studies in the context of health to support medical decision-making.

A systematic review was first conducted to document segmentation methods used in the health-focused stated-preference studies in current literature. It identified current practices and knowledge gaps. LCA is then applied to empirical stated-preference data generated by two most commonly used stated-preference methods identified in the systematic review, namely discrete choice experiment (DCE) and best-worst scaling (BWS). Model specifications were modified in both applications to better serve policy and clinical decision-making.

Latent class logit (LCL) is the most commonly used segmentation model that has been applied in both DCE and BWS. However, both applications have limitations. LCL sometimes over-fits the DCE data and leads to too many classes that are difficult to incorporate in policy or clinical decision-making when there is substantial within-class preference heterogeneity or significant overlap between classes. Random effects are incorporated in LCL models as a remedy in this dissertation. With more flexible model specification, random effect LCL is shown not only reducing the number of classes but also better capturing the complex and dispersed preference pattern among patients than LCL, leading to improved model fit and prediction accuracy.

When LCL is applied to BWS data, information criteria also often fail to identify the best-fitted model and parsimonious segmentation results due to cross-task constant utility assumption in LCL. A standard LC model was used to relax the constraint in this dissertation. It dramatically reduced the number of classes and generated practical segmentation results. Given that regulatory and clinical decision-makers often prefer parsimonious results due to limited resources and capacity to accommodate too many preference types, more flexible model specifications should be used in stated-preference studies to generate more practical results to support decision-making.

Readers: Dr. John F.P. Bridges (advisor), Dr. Karen Bandeen-Roche, Dr. Darrel Gaskin, Dr. Jodi Segal

Preface/Acknowledgement

This dissertation received funding from the Patient-Centered Outcomes Research Institute Methods Program Award (ME-1303-5946). I am grateful for the guidance from my academic advisor Dr. John F.P. Bridges as well as my thesis committee members Dr. Karen Bandeen-Roche and Dr. Jodi Segal. I appreciate the knowledge and expertise they shared with me. It has been a pleasure working with them. The dedication to research and scientific integrity they showed me helped me grow academically.

I would also like to thank my parents, family, and friends for the support and love they gave me throughout my life. I especially want to acknowledge my mother, who passed away one year before my defense. However, it is her guidance, support, and what she taught me throughout my life that got me to where I am today. I would not be able to succeed without them.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

Patients' preferences over medical treatments are essential because only patients live with their medical conditions and the consequence of treatment choices. Their perspective can be different from the perspective of regulators and healthcare providers. As researchers have been increasingly studying patient preferences to inform policy and clinical care, most research focuses on the average preference among patient population (Clark et al., 2014). However, significant preference heterogeneity may exist among patients due to different socioeconomic background, culture, experience, believe, or personality. For example, low-income people are probably more concerned about the cost of a treatment relative to other characteristics than higher-income people. Even with the same socioeconomic status, two individuals may still have different preferences due to different experience. When there is preference heterogeneity, the policy and clinical decisions made based on the average preference among patients may lead to low satisfaction and low treatment adherence or uptake rate.

To further illustrate, suppose we are studying the relative importance of treatment cost and effectiveness in a group of patients. Suppose there are two major types of preferences in the population. As shown in Figure 1-1, where each axis represents the utility for each attribute, one group of individuals (the black dots) values cost more than effectiveness, while the other group (the blue dots) cares about effectiveness more than cost. The average preference among the entire population is where the vertical and horizontal lines cross. If we develop and price treatment based on this average preference, a majority of the cost-focused group will be unwilling or unable to pay for the treatment, whereas most of the effectiveness-focused group will be unsatisfied with the effectiveness of the treatment and wiling to pay more for more effective alternatives.

Knowledge about patient preference heterogeneity is valuable for policy and clinical decision-making when there is significant heterogeneity in the population. Market segmentation and product differentiation based on preference heterogeneity will increase the overall patient utility (Smith, 1956). Using the above example, cheaper treatments should be provided to the cost-conscious individuals, while more effective alternatives, if any, should be accessible in the effectiveness-focused group even if a higher price is required to cover the cost. Such tailored treatment plans will improve treatment adherence, which leads to better health outcomes (Capoccia et al., 2016). Although limited resources may not allow policy-makers and clinicians to accommodate each individual preference, different treatments or treatment plans should be provided to meet the significantly distinct needs.

1.1 Conceptual Model to Study Preference

The conceptual model on preference in Figure 1-2 gives us some hint on how we can study preference heterogeneity. The model is adapted from Simonson's conceptual framework on determinants of customers' responses to customized offers (Simonson, 2005). As preference is difficult to directly observe, it is often inferred through choices. When people make choices on a product (or service), they observe the attributes (or features) of the product and evaluate them based on their preference to determine if the product is attractive. If the attributes fit the decision maker's preference, the product is perceived as attractive and is likely to be chosen. If the attributes do not fit the decision maker's preference, the product is unlikely to be chosen.

Evaluation of the attributes can be affected by several factors. Individuals often compare options within a choice set and give little considerations to those outside of the choice set, especially when they do not have clear and strong preference. Choice context, such as the choice scenario and the options available in the choice set (i.e., what they are comparing with), therefore influences the assessment of each product. Presentation format and the number of options available in the choice set can also have a significant impact on responses. For example, a large number of options decrease the likelihood of any single option being perceived as sufficiently attractive. Presenting the options in different order may also lead to different responses. (Simonson, 2005)

This dissertation focuses on analyzing preference data from stated-preference surveys, the methods that have been increasingly applied to valuate healthcare services from patients' perspective in the past two decades (de Bekker-Grob et al, 2012; Clark et al, 2014). Different from revealed preference data where real-world choices are observed, stated-preference methods use survey to ask people to make choices under hypothetical scenarios. Product attributes, presentation context, and presentation format are therefore controlled in an experimental environment. As a result, stated-preference methods allow us to reveal patients' true preference without confounding from other factors such as physician preference, insurance status, and treatment availability as in patients' real-world treatment choices. Heterogeneous choices among respondents in a stated-preference study can then be linked to heterogeneous preferences, after controlling for random errors. Despite the hypothetical choices, stated-preference methods have been shown to be valid predictors of consumers' actions (Louviere et al, 1981; Kocur et al, 1982).

Preference for medical treatments and health services can be affected by various factors. Some such as demographics, socioeconomic status, and disease history, can be easily observed. For example, preference for diabetes medication has been shown to differ by age (Hauber et al, 2016), gender (Gelhorn et al, 2013), and previous experience with injectable diabetes medicine (von Arx et al., 2014). Other factors that influence preference may not be easy to measure or document. For example, the variation in attitudes and behaviors about diabetes management across patients has been shown to reflect individual's knowledge and opinions rather than patient's age, gender, or culture (Onwudiwe et al., 2011).

Researchers traditionally use stratification to compare the preference of sub-groups. They separate study subjects into homogeneous groups based on observed characteristics (e.g., demographics) and estimate either separate models or separate sets of coefficients for each strata (Adamowicz et al, 1997). Stratification assumes that preference heterogeneity can be accurately determined *a priori* by observed variables (Boxall et al, 2003), which, as shown above, often does not hold empirically (Morey and Greer Rossmann, 2003; Iraguen and de Dios Ortuzar, 2004). When observed and unobserved factors jointly influence preference, stratification on a limited number of observables is insufficient to explain the variation in preference. In addition, even when the assumption holds, preference heterogeneity can be misclassified if stratification is performed based on the wrong characteristics. For example, preference variation may be identified across levels of educational attainment, but they may in fact relate to differences in the respondents' income.

An emerging alternative to stratification is segmentation, where respondents are classified into groups or clusters based on the patterns of choices or preference (Cunningham et al, 2008; Hole, 2008). Segmentation assumes that preferences are distributed discretely into like clusters. They are not simply a function of demographic variables, but are formed by perceptions, experience, beliefs and unobserved variables (Hilger and Hanemann, 2006). Segmentation is usually estimated simultaneously with the choice model to identify population subsets composed of like-minded individuals with homogeneous preference (McFadden, 1986). Since segmentation is probabilistic (i.e. respondents are allocated to the group that they are most likely to be a member of), multivariate statistics can be used to analyze what individual characteristics are in fact correlated with class membership, controlling for potential confounding factors. The following chapters in this dissertation focus on applying segmentation methods, especially latent class analysis, to analyze preference heterogeneity in stated-preference studies.

1.2 Statistical Model to Analyze Preference

Random utility theory (RUT) is often applied in the analysis of choice data (McFadden, 1974). The theory is grounded in both psychology and economics, and provides a rigorous and relevant approach to understand people's preference and choices (Thurstone, 1927). According to the theory, the utility an individual n obtains from choosing an object i can be described as a function, $U_{ni}(.)$, of the observed attributes of object i, V_{ni}, and the unobserved characteristics and disturbances, expressed as a random error term ε_{ni}:

$$U_{ni} = V_{ni} + \varepsilon_{ni} \qquad [1]$$

When there are only two objects in a choice set, i and j, i j = 1,2, the probability that i is chosen over j is the probability that the utility person n obtains from i is larger than the utility from j, that is:

$$P_{ni} = \Pr\left(U_{ni} > U_{nj}\right) = \Pr\left(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj}\right) = \Pr\left(\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} - V_{nj}\right) \quad [2]$$

If the error terms, $\varepsilon_{n\bullet}$, are independently and identically distributed (IID) and follow a type I extreme value distribution (also called Gumbel distribution), the difference of the error terms in equation [2], ε_{nj} - ε_{ni}, follows a logistic distribution (McFadden, 1974), and P_{ni} has a closed form:

$$P_{ni} = \frac{\exp\left(V_{ni}\right)}{\exp\left(V_{ni}\right) + \exp\left(V_{nj}\right)} \quad [3]$$

When there are more than two options in the choice set, the probability that person n chooses option i, $i = 1,...,J$, among J alternatives is the probability that the utility person n obtains from i is larger than the utility from any other alternatives in the choice set, that is:

$$P_{ni} = \Pr\left(U_{ni} > U_{nj}\right) = \Pr\left(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj}\right) = \Pr\left(\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} - V_{nj}\right) \quad [4]$$

for all $j \neq i$ and $j = 1,\ldots,J$. If the error terms are IID and follow a type I extreme value distribution, P_{ni} has a closed form and can be expressed as:

$$P_{ni} = \frac{\exp\left(V_{ni}\right)}{\sum_{j=1}^{J} \exp\left(V_{nj}\right)} \quad [5]$$

Given the probabilities, the odds that option i is chosen over option j, $i, j = 1,\ldots J$, is:

$$\frac{P_{ni}}{P_{nj}} = \frac{\exp\left(V_{ni}\right)}{\exp\left(V_{nj}\right)} = \exp\left(V_{ni} - V_{nj}\right) \quad [6]$$

Representative utility is usually specified to be linear in parameters, i.e. $V_{ni} = x'_{ni}\beta$, where x_{ni} is a vector of attributes relating to object i and β is a vector of utility values for these attributes. With this specification, if we take the logarithm of equation [6] and substitute V_{ni} and V_{nj} with object attributes, we can simply estimate:

$$\ln\left(\frac{P_{ni}}{P_{nj}}\right)=V_{ni}-V_{nj}=x'_{ni}\beta-x'_{nj}\beta=\left(x'_{ni}-x'_{nj}\right)\beta \quad [7]$$

which is the conditional logit regression that has been widely used to analyze discrete choice data in health (Hauber et al., 2016).

In segmentation models, we assume β follows a discrete distribution and only takes Q distinct values. A latent class logit model (LCL, also termed finite mixture logit model) is often applied to discrete choice data to perform market segmentation (McFadden, 1986; Swait, 1994). It consists of a measurement model estimating class-specific preferences and a conditional structural model determining class membership based on choice patterns and individual characteristics. Suppose an individual n belongs to a market segment q, $q = 1,\ldots,Q$. The measurement model for the conditional probability that person n in segment q chooses i among J alternatives is:

$$P_{ni|q}=P_n\left(i|q\right)=\frac{\exp\left(x'_{ni}\beta_q\right)}{\sum_{j=1}^{J}\exp\left(x'_{nj}\beta_q\right)}. \quad [8]$$

Preference heterogeneity is captured by the segment-specific utility parameter β_q.

The structural model on the other hand assumes that membership within a market segment is a function of individual characteristics (e.g. demographics), Z_n, and an error term. Again, assuming the error term is type I extreme value distributed, the probability that person n belongs to segment q can be specified via a multinomial link function of individual covariates:

$$\pi_{nq}=P_n\left(q\right)=\frac{\exp\left(z'_n\delta_q\right)}{\sum_{q=1}^{Q}\exp\left(z'_n\delta_q\right)} \quad [9]$$

where δ's are the structural parameters to be estimated. For model identification, one of the Q segments (e.g. $q=Q$) is typically chosen as the reference segment by setting $\delta_Q=0$. Then δ_q measures change in the odds of belonging to segment q relative to the reference segment Q due to each one-unit change in Z_n.

Estimation of the LCL model is conducted by joining equations [8] and [9] under the non-differential measurement assumption. Empirically, the probability of person n in segment q choosing the object that he is actually observed to choose can be expressed as $\prod_{i} \left(P_{ni|q}\right)^{y_{ni}}$, where $y_{ni}=1$ if person n chose i and zero otherwise. Since $y_{ni}=0$ for all non-chosen objects and $P_{ni|q}$ raises to the power of zero is 1, this term is simply the probability of the chosen object. Respondents often make choices in a series of scenarios in stated-preference studies. Suppose there are T choice tasks in a survey. The probability that person n in segment q makes a sequence of choices as observed can be expressed as $\prod_{t=1}^{T} \prod_{i} \left(P_{nit|q}\right)^{y_{nit}}$, assuming conditional independence (Hole, 2008). When preference heterogeneity is taken into consideration, the probability of a sequence of choices being observed from person n is:

$$S_n = \sum_{q=1}^{Q} \pi_{nq} \prod_{t=1}^{T} \prod_{i} \left(P_{nit|q}\right)^{y_{nit}} \qquad [10]$$

Assume choices are independent among decision makers. Then the log-likelihood function is:

$$LL(\beta) = \sum_{n=1}^{N} \ln \left\{ \sum_{q=1}^{Q} \pi_{nq} \prod_{t=1}^{T} \prod_{i} \left(P_{nit|q}\right)^{y_{nit}} \right\} \qquad [11]$$

where β is a vector containing all parameters of the model. The estimator is the value of β that maximizes this function. Because it is impossible to directly solve the equations from the log-likelihood function, Expectation-Maximization and Newton-Raphson algorithms are often used to estimate the maximum likelihood solution (Huang and Bandeen-Roche, 2004).

In the following chapters, I first overview how latent class analysis has been used to study preference heterogeneity in stated-preference studies in health. Current practice is summarized based on a systematic review and compared to the analytical recommendation to date. The LCL model is then applied to the empirical stated-preference data generated by two most commonly used stated-preference methods identified in the systematic review, namely discrete choice experiment (DCE) and best-worst scaling (BWS). I examine and improve the application by modifying model specification to better serve policy and clinical decision-making.

Reference

Adamowicz W, Swait J, Boxall P, et al. Perceptions versus objective measures of environmental quality in combined revealed and stated preference models of environmental valuation. Journal of Environmental Economics and Management. 1997;32(1):65-84.

Boxall PC, Englin J, and Adamowicz WL. Valuing Aboriginal artifacts: A combined revealed-stated preference approach. Journal of Environmental Economics and Management. 2003;45(2):213-30.

Capoccia K, Odegard PS, and Letassy N. Medication adherence with diabetes medication: a systematic review of the literature. The Diabetes Educator. 2016;42(1):34-71.

Clark MD, Determann D, Petrou S, et al. Discrete choice experiments in health economics: a review of the literature. Pharmacoeconomics. 2014;32(9):883-902.

Cunningham CE, Deal K, Rimas H, et al. Using conjoint analysis to model the preferences of different patient segments for attributes of patient-centered care. The Patient: Patient-Centered Outcomes Research. 2008;1(4):317-30.

de Bekker-Grob EW, Ryan M, and Gerard K. Discrete choice experiments in health economics: a review of the literature. Health Econ. 2012;21(2):145-72.

Gelhorn, H.L., Stringer, S.M., Brooks, A, et al. Preferences for medication attributes among patients with type 2 diabetes mellitus in the UK. Diabetes Obes Metab. 2013;15(9):802-9.

Hauber AB, Gonzalez JM, Groothuis-Oudshoorn CG, et al. Statistical methods for the analysis of discrete choice experiments: a report of the ISPOR conjoint analysis good research practices task force. Value in Health. 2016;19(4):300-15.

Hauber AB, Nguyen H, Posner J, et al. A discrete-choice experiment to quantify patient preferences for frequency of glucagon-like peptide-1 receptor agonist injections in the treatment of type 2 diabetes. Curr Med Res Opin. 2016;32(2):251-62.

Hilger J and Hanemann M. Heterogeneous preferences for water quality: A finite mixture model of beach recreation in Southern California. San Diego: Research Final Reports, California Sea Grant College Program, UC San Diego; 2006.

Hole AR. Modelling heterogeneity in patients' preferences for the attributes of a general practitioner appointment. Journal of Health Economics. 2008;27(4):1078-94.

Huang GH, and Bandeen-Roche K. Building an identifiable latent class model with covariate effects on underlying and measured variables. Psychometrika. 2004;69:5–32.

Iraguen P, and de Dios Ortuzar J. Willingness-to-pay for reducing fatal accident risk in urban areas: an internet-based web page stated preference survey. Accident Analysis & Prevention. 2004;36(4):513-24.

Kocur G, Hyman W, Aunet B. Wisconsin work mode-choice models based on functional measurement and disaggregate behavioral data. Transportation Research Record. 1982;895:24-32.

Louviere JJ, Henley D, Woodworth G, et al. Laboratory-simulation versus revealed-preference methods for estimating travel demand models. Transportation Research Record. 1981;794:42-51.

McFadden D. Conditional logit analysis of qualitative choice behavior. In: P. Zarembka, editor. Frontiers in econometrics. New York: Academic Press; 1974.

McFadden D. The choice theory approach to market research. Marketing Science. 1986;5(4):275-97.

Morey E and Greer Rossmann K. Using stated-preference questions to investigate variations in willingness to pay for preserving marble monuments: Classic heterogeneity, random parameters, and mixture models. Journal of Cultural Economics. 2003;27(3):215-29.

Onwudiwe NC, Mullins CD, Winston RA, et al. Barriers to self-management of diabetes: a qualitative study among low-income minority diabetics. Ethnicity & Disease. 2011;21(1): 27-32.

Simonson I. Determinants of customers' responses to customized offers: conceptual framework and research propositions. Journal of Marketing. 2005;69:32-45.

Smith WR. Product differentiation and market segmentation as alternative marketing strategies. The Journal of Marketing. 1956;21(1):3-8.

Swait J. A structural equation model of latent segmentation and product choice for cross-sectional revealed preference choice data. Journal of Retailing and Consumer Services. 1994;1(2):77-89.

Thurstone LL. A law of comparative judgment. Psychological Review. 1927;34(4):273-86.

von Arx LB, and Kjeer T. The patient perspective of diabetes care: a systematic review of stated preference research. Patient. 2014;7(3):280-300.

Figure 1-1 Preference heterogeneity among patients

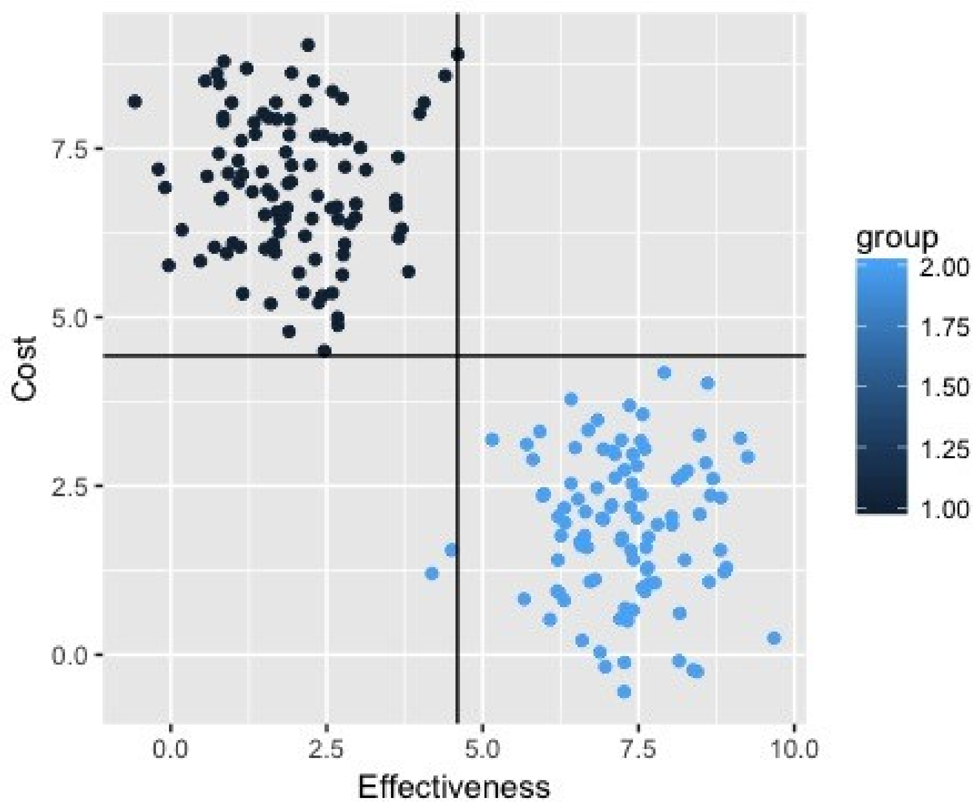
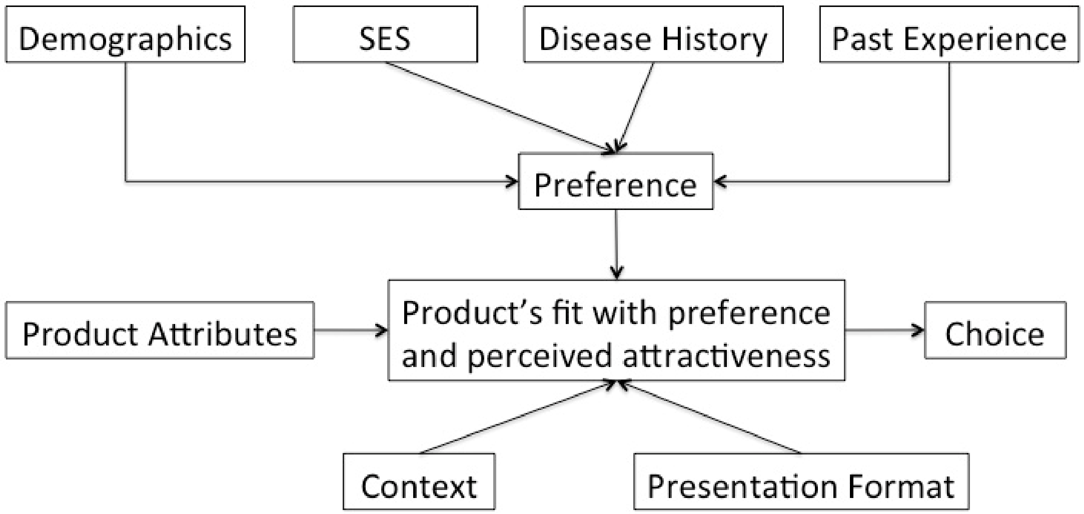


Figure 1-2 Conceptual model for preference



Chapter 2 Using Latent Class Analysis To Model Preference Heterogeneity In Health: A Systematic Review

Mo Zhou, MPA, MHS; Winter Maxwell Thayer; John F P Bridges, PhD

Abstract

Background: The preferences of patients and other stakeholders in health are increasingly being used to inform decision-making. Latent class analysis (LCA) is increasingly used to document and explore preference heterogeneity. Using LCA is both an art and science, yet there is little guidance or standards on applying it in health.

Objective: We sought to document the applications of LCA in the health-focused stated-preference literature and to inform future studies by identifying current norms in published applications.

Methods: We conducted a systematic review of the MEDLINE, Embase, EconLit, Web of Science, and PsycINFO databases. We included stated-preference studies that used LCA to explore preference heterogeneity in healthcare or public health. Two reviewers independently evaluated titles, abstracts, and full-text articles. Key outcomes abstracted included segmentation methods, preference elicitation methods, number of attributes and levels, sample size, model selection criteria, number of classes reported, and hypotheses tests. Study data quality and validity was assessed with the PREFS (purpose, respondents, explanation, findings, significance) quality checklist.

Results: Among the 2,560 identified titles, 99 met the inclusion criteria for the review. Most studies focused on the preference of patients (34.3%) and the general population (31.3%). Nearly 80% used discrete choice experiments. The number of attributes included in the studies ranged from 3 to 20, with over half of the studies having four to six attributes. Sample size in LCAs ranged from 47 to 2,068, with a third between 100 and 300. Over 90% of the studies used latent class logit models for segmentation. Bayesian information criterion (BIC, 59.0%), Akaike information criterion (AIC, 53.8%), and log-likelihood (LL, 28.2%) were commonly used for model selection, while class size and interpretability were also considered in some studies. About 80% of studies reported two to three classes. Only 30% used statistical tests to detect significant variation in preference between classes. Individual characteristics were included in segmentation model in 47.5% of the studies, while 30.3% conducted post-estimation analyses to examine class characteristics. 36% and 43% of studies discussed implications for policy or clinical practice.

Conclusions: LCA is increasingly used to study preference heterogeneity in health and support decision-making. As its application in health is relatively new, there is little consensus on best practices. Increasing emphasis on patient preferences from regulatory agencies and patient-centered care in clinical settings is likely to lead to greater demand for methods to quantify preference heterogeneity. Guidance is needed to improve the quality of LCA studies in health to support policy development and clinical practice.

2.1 Introduction

Wide acceptance of the principles of patient-centered care has led to an increase in interest in obtaining patient input on regulatory evaluations (Califf, 2017). The Food and Drug Administration (FDA) recently issued guidance for premarket reviews and approval that listed patient preference information as an important consideration for regulatory reviews (CDRH, 2015). The FDA recognizes heterogeneity among patients in risk tolerance and benefit-risk tradeoff and has expressed willingness to approve medical devices with a benefit-risk profile that is only acceptable to a subset of patients (CDRH, 2015). Understanding patient preferences and preference heterogeneity can support patient-centered regulatory decision-making and expand access to effective treatments.

Stated-preference methods have been increasingly employed to valuate healthcare services from patients' perspective over the past two decades (de Bekker-Grob et al., 2012; Clark et al., 2014). Stated preference methods measure utility in order to quantify the strength of individuals' preferences (Bateman et al., 2002). In choice-based conjoint analysis (e.g., discrete choice experiments) health products or health services are decomposed into attributes with varying levels. Researchers can evaluate respondents' choices between competing options to understand which attributes they value most.

Less than half of stated-preference applications in health explored preference heterogeneity across patient subgroups despite increased use of the methods (Clark et al, 2014). Heterogeneity has traditionally been explored by stratifying participants into homogeneous groups based on observed characteristics (e.g., demographics) (Adamowicz et al., 1997). However, stratification requires a limiting assumption that preference heterogeneity can be accurately determined *a priori* by observed variables

(Boxall et al., 2003), an assumption that sometimes does not hold empirically (Morey and Greer Rossmann, 2003; Iraguen and de Dios Ortuzar, 2004).

Segmentation is an emerging alternative to stratification that classifies respondents into groups or clusters based on response patterns (Deal, 2014). Segmentation is estimated simultaneously with the choice model to identify subsets of participants with preferences that are homogeneous within groups but heterogeneous between groups (McFadden, 1986). Individuals' preferences are assumed to be formed by perceptions, experience, beliefs and unobserved variables, rather than simply a function of observed variables (Hilger and Hanemann, 2006). Segmentation is probabilistic (i.e., respondents are allocated to the group that they are most likely to be a member of), and multivariate statistics can be used to describe differences in characteristics across groups.

Few studies adequately describe the application of these statistical methods in preference heterogeneity in health. This review aims to document how segmentation methods such as latent class analysis (LCA) have been used to study preference heterogeneity among stated-preference studies conducted in the health field. Best practices and methodological gaps in current literature are identified to provide a methodological guide for researchers who study preference heterogeneity in health.

2.2 Methods

2.2.1 Search Strategy

We performed a literature search in MEDLINE, Embase, EconLit, Web of Science, and PsycINFO databases. Search terms consisted of key words for stated-preference methods, including 'stated preference', 'conjoint analysis', 'conjoint analyses', 'discrete choice', 'choice experiment', 'best worst scaling', 'choice survey', and ‘contingent valuation’, as well as segmentation methods, including ‘latent class’, ‘finite mixture’, ‘segmentation’, ‘heterogeneity’, ‘heterogeneous’, and ‘hierarchical bayes’. The final search was conducted on December 31st 2016 and includes all publications up to the end of 2016.

2.2.2 Inclusion and Exclusion Criteria

Papers were included if they 1) used stated-preference methods to elicit preference, 2) were related to healthcare or public health, 3) analyzed preference heterogeneity using segmentation, and 4) were published in peer-reviewed journals. Articles were excluded if they 1) were not written in English, 2) did not use stated-preference methods, 3) were not related to healthcare or public health, 4) did not analyze preference heterogeneity using segmentation, 5) contained no original data (e.g., review, commentary, editorial, book chapter, or meeting abstract), or 6) were not peer-reviewed (e.g., working papers, dissertation, or grey literature). Articles related to food choices (e.g., organic food, genetically modified foods) and environmental policies (e.g., air quality control) were excluded.

2.2.3 Data extraction and Analysis

Two reviewers (MZ and WT) independently screened titles then abstracts. Titles and abstracts were excluded if both reviewers determined that they did not meet the inclusion criteria. The two reviewers then independently reviewed the full text of the remaining articles. Disagreements about inclusion and exclusion based on full-text review were determined by consensus between the two reviewers and in case of remaining doubt, by a third researcher (JB).

For each included article, the two reviewers abstracted data independently using a pre-defined data abstraction form. Key outcomes comprised segmentation methods, preference elicitation methods, number of attributes and levels, sample size, model selection criteria, number of classes reported, and hypotheses tests. Study topic, year, country of origin, study population, alternative models used in the study in addition to segmentation model, prediction approaches, policy relevance, and software used for segmentation were also abstracted. Ordinary least squares (OLS) was used to explore factors that were associated with the key outcomes such as the number of attributes and levels, sample size, and the number of classes reported. Chi-square tests and *t*-tests were used to detect significant differences between studies.

2.3 Search Results

Among the 2,560 identified citations, 186 articles were eligible for full-text review. 99 studies met our inclusion criteria and were included in the final review (Fig. 2-1). Among the excluded articles, 31.3% did not use stated-preference methods to measure preference, 28.1% were not related to healthcare or public health[1], 22.6% did not original data, and 17.2% did not use segmentation methods to explore heterogeneity.

[FIGURE 2-1 INSERT HERE]

2.3.1 Trend in Time and Geographic Location

Fig. 2-2 shows the number of studies by year. The earliest stated-preference study that used segmentation methods to explore preference heterogeneity in the health field was published in 1996. Only four studies were published between 1996 and 2006. Segmentation became more widely used after 2011, with the number of publications

[1] Some of these studies examined preference for food items (e.g., organic food, genetic modified food) that have undetermined health impacts.

increasing from seven in 2011 to 27 in 2016. A majority (87.9%) of the studies in our review were published between 2011 and 2016.

[FIGURE 2-2 INSERT HERE]

Researchers in the United States (US) were the first to apply segmentation methods in health-related stated-preference studies to study preference heterogeneity. Before 2012, only Canada, US and the United Kingdom used segmentation methods in their preference research. The application has been expanded to more countries in the past five years, with the number of countries increasing from three in 2012 to 12 in 2016. Overall, most studies were conducted in Canada (29.3%), the US (15.1%), the UK (12.1%), the Netherlands (11.1%), and Australia (10.1%). Studies also took place in other developed countries (e.g., Germany, Italy, Switzerland, France, Spain, and Singapore) and some developing countries (e.g., China, Ghana, South Africa, Uganda, and Malawi). Two studies were performed across several countries in Europe.

2.3.2 Study Topic and Population

The most common study topic was preference for medical treatments (40.4%), followed by preference for prevention programs of physical (e.g., obesity, injury) or mental illnesses (e.g., depression) (22.2%). Three studies concerned preferences for screening tests. The most common disease categories include mental illnesses (11), cardiovascular diseases (5), and cancer (5). Twelve studies (12.1%) evaluated the value of primary, personalized, or patient-centered health services in general, and eight (8.1%) assessed quality of life rather than focusing on a specific disease. In addition to healthcare, studies also touched on job and training preferences among health care providers or medical students (7), health care resource allocation (5), and the preference of organ donation policies (1).

A majority of the studies were conducted among patients (34.3%), the general population (31.3%), and informal caregivers such as parents (18.2%). Only 12 studies evaluated clinicians' preference, while some investigated the preference of other health services providers (2), clinic staff (2), and pharmacists (1). Two studies elicited policymakers' opinions about health care resource allocation. Three articles surveyed medical or nursing students for job and training preferences. Research was also conducted among educators (2) and students (3) to evaluate school-based prevention programs.

Eleven studies surveyed more than one population. Patients were included in nine, and their preferences were compared to the caregivers (6), clinicians (4), or the general population without the health condition (1). Two papers surveyed three study populations (i.e., patients, caregivers, and clinicians) in one study.

One-third of the studies restricted respondents to adults (i.e., >18 years old). Five articles assessed preferences among children, with age range between 9 and 17 years old, and three articles evaluated preferences among seniors (i.e., ≥ 65 years old). Most (61.6%) of the studies did not have any age restrictions.

2.3.3 Preference Elicitation Methods

Majority of the studies (78.8%) used discrete choice experiment (DCE) to elicit preferences. Seventeen studies employed best-worst scaling (BWS). Segmentation of BWS data was first conducted in 2010 and has been increasing since 2015. Compared to DCE, segmentation of BWS data was less common in terms of the total number of applications and geographic locations (16 countries for DCE vs. 7 countries for BWS). Four articles used rating to evaluate the value of health service profiles or treatment outcomes.

The number of attributes included in the studies ranged from 3 to 20, with a mean of 7.8 and a median of 6. (Fig. 2-3) Over half of the studies (56.6%) had four to six attributes. A majority of the studies with 10 or more attributes were conducted in Canada (16) or the US (6). When there were more than 10 attributes, choice tasks often only included a subset of attributes to reduce respondent burden and respondents were randomly assigned to answer one version of choice tasks. The number of attribute in BWS studies ranged from 5 to 16.

[FIGURE 2-3 INSERT HERE]

A linear regression (Table 2-1) showed that on average, studies conducted in Canada had 3.1 more attributes than those from the US ($p = 0.025$), holding study population, age group, and stated-preference methods constant. Surveys among patients and general population included 2.5 ($p = 0.013$) and 3.3 ($p = 0.004$) fewer attributes, respectively, compared to the rest, probably to reduce respondent burden. There was no difference in the number of attributes between different age groups or different stated-preference methods ($p > 0.05$). No significant time trend was found regarding the number of attributes included in the studies ($p = 0.101$).

[TABLE 2-1 INSERT HERE]

The number of levels ranged from one to eight. Most (82.8%) of the studies had no more than four levels for each attribute. A multivariate analysis suggested that the maximum number of levels included in a study was not correlated with the number of attributes or study population, but varied geographically and between different stated-preference methods (Table 2-1). Studies from Canada and the UK had 1.3 ($p = 0.001$) and 0.9 ($p = 0.051$) more levels, respectively, than those from the US. BWS had 0.7 fewer levels on average than DCEs ($p = 0.033$) because BWS object case or case 1 only allows one level for each attribute. Nearly half of the studies (51.0%) had unbalanced numbers of levels. A logistic regression showed that one additional attribute in a study was associated with a 4.5% increase in probability of having balanced design ($p = 0.005$), for a given stated-preference method, study population, and country of origin. This may suggest that researchers try to simplify study design when the number of attributes is larger.

2.3.4 Sample Size

Overall, sample size per study population ranged from 19 to 2,068, with a mean of 479 and a median of 290. Sample size in LCAs was relatively larger, ranging from 47 to 2,068, with a mean of 497 and a median of 303, because several studies merged choice data from several study populations (e.g., patients and caregivers) in the segmentation analysis when one group was too small (Fig. 2-4). One-third of LCAs had sample sizes between 100 and 300, while fewer LCAs were performed on a sample below 100 (12.8%) or above 1,000 (22.9%).

[FIGURE 2-4 INSERT HERE]

No time trend was found regarding sample size in LCAs ($p = 0.133$). A multivariate analysis indicated that sample size was associated with the number of attributes, country of origin, and study population (Table 2-2). One additional attribute was correlated with an average of 58 more respondents ($p < 0.001$). Sample sizes from the US were relatively smaller than other regions, but only those from Australia were significantly higher (by 548) (p = 0.014). Surveys among the general population had 360 more respondents on average than those among patients, providers, or caregivers (p = 0.008), possibly due to larger sampling pools.

[TABLE 2-2 INSERT HERE]

2.4 Segmentation Procedures

Most (92.9%) studies used latent class logit models for segmentation, nine of which considered heterogeneity in scale, and one study (Deal, 2014) compared probabilistic LCL model to the distance-minimizing mathematical procedures such as clues (CLUstEring based on local Shrinking) and pamk (Partitioning Around Medoids). Seven studies used k-mean cluster analysis for segmentation. Studies with rating techniques exclusively used k-mean cluster analysis for segmentation, whereas LCL has only been used to analyze discrete choice data (i.e. DCE or BWS).

2.4.1 Model Selection Criteria

Overall, 81.8% of the articles listed the criteria they used to identify the appropriate number of classes. A majority of the k-mean cluster analyses did not specify the criteria. Among the three articles (out of seven) that listed the model selection criteria, they used the Ward’s method, maximized Kappa statistics, and externally validated the relative importance against the self-reported values to determine the number of classes. Researchers also considered the interpretability of the classes to decide whether to add or drop a class in a model.

Most studies (84.8%) that used LCL model specified model selection criteria, among which 66.7% used more than one criterion to identify the number of classes. The most common criteria included Bayesian information criterion (BIC, 59.0%), Akaike information criterion (AIC, 53.8%), and log likelihood (LL, 28.2%), with 85.9% of the studies using at least one of these criteria. Other information criteria such as consistent AIC (CAIC, 12.8%), Hannan-Quinn information criterion (HQIC, 2.6%), and AIC3 (1.3%) were also used in some studies. Statistic indicators such as Entropy R-square, McFadden's R-square (or rho-square), and Swchwarz criterion were sometimes examined in addition to information criteria to help determine the optimal model. Some researchers went beyond the statistics and considered interpretability of the classes (16.7%), segment size (11.5%), or the variation between classes (2.6%) while evaluating model fit. Two studies predicted choices in the holdout tasks using preference estimates and chose the best-fitted model based on the choice hit rate. One study determined the number of classes based on theory.

2.4.2 Number of Classes

Most studies (95) reported the optimal number of classes. Among studies that did not report the number of classes, two stated that the aggregated model fit the data better than segmentation model and did not report the segmentation results; one study estimated solutions with two to five classes without choosing the best-fitted model; and one study used LCA to examine attribute non-attendance and did not focus on finding the optimal number of classes.

The reported number of classes ranged from two to seven, with a mean of 2.8 and a median of 3. 80.2% of studies reported two to three classes, with 2-class model selected more often than 3-class model. Only five studies reported more than four classes, with four studies reporting five classes and one study reporting seven classes. A scatter plot between the number of classes reported in the study and sample size suggested a positive correlation between the two (Fig. 2-5).

[FIGURE 2-5 INSERT HERE]

A linear regression showed that only sample size was correlated with the number of classes reported. Every thousand respondents were associated with a 0.5 increase in the number of classes, holding other study characteristics constant ($p = 0.046$). The number of classes did not vary systematically between different numbers of attributes or levels, country of origin, study population, stated-preference methods, segmentation methods, or model selection criteria ($p > 0.05$) (Table 2-3).

[TABLE 2-3 INSERT HERE]

2.4.3 Comparison of Preferences and Characteristics

Most (69.7%) of studies examined preference variation between classes by simply comparing the magnitude of preference weights or relative importance without conducting any statistical tests. Among the 30 articles that used statistical tests to detect significant differences between classes, 21 were from Canada. 15 studies (50%) used Wald test, which had been widely used since 2011. Other statistic tests for cross-class comparison of preferences included MANCOVA, ANCOVA, Dunnett's C test, Tukey comparison, and *t*-test, which were predominantly used by researchers in Canada.

Almost half of studies (47.5%) included respondent characteristics in the segmentation model to facilitate segmentation. Seven studies (14.9%) used AIC, BIC, or LL ratio test to determine which characteristics should be included. Twenty-eight of the 47 studies (59.6%) identified individual characteristics associated with class membership based on the coefficients of the characteristic variables in the segmentation model. Chi-square test, ANOVA, MANOVA, and logistic regressions were used additionally in several studies to analyze how respondent characteristics varied between different classes.

One-third of studies (30.3%) did not include individual characteristics in the segmentation model but conducted post-estimation analyses to examine how individual characteristics differ between classes. Depending on the distribution of characteristic variables, chi-square test (56.7%), logistic regression (40.0%), ANOVA or MANOVA (26.7%), t-test (6.7%), or Wilcoxon rank sum test (3.3%) was used to test the correlation between class membership and characteristics. Discriminant analysis and random forests were also conducted to analyze class characteristics.

2.4.4 Alternative Models and Prediction

Among 78 DCE studies, 34.6% used only segmentation models to analyze preference; 46.2% compared segmentation model to the aggregated conditional logit or multinomial logit model; and 11.5% used Hierarchical Bayes to estimate individual preference weights. (Fig. 2-6) Although mixed logit (MXL) has been widely used to analyze preference heterogeneity in recent years, only 12 studies (15.4%) compared LCA results to MXL models. Logit, generalized multinomial logit (GMNL), nested logit, and linear probability models (LPM) were also employed in some studies. Over half of the studies (54.9%) with alternative models statistically compared the performance between different models. Common criteria included LL (29.4%), AIC (27.4%), BIC (17.6%), McFadden's adjusted R-square (7.8%), and the percentage of correctly predicted choices (5.9%).

[FIGURE 2-6 INSERT HERE]

Among 17 BWS studies, 17.6% used only segmentation models to analyze preferences. Less than half of studies (41.2%) compared the segmentation model to the aggregated conditional logit or multinomial logit model. Two studies (11.8%) estimated individual preference weights using Hierarchical Bayes, and only one study (5.9%) included MXL model. (Fig. 2-6) Best-worst scores, MaxDiff, and rank-ordered logit have also been used to analyze BWS data. Only one study used statistical methods to compare performance between models.

One-third of studies (34.3%) predicted respondent choices and market uptake using estimated preference weights from the segmentation model. Over half (55.9%) of these studies were from Canada. In terms of prediction methods, 58.8% of studies performed randomized first-choice simulation. 11 studies (32.4%) conducted scenario analyses and estimated the expected uptake of given treatment or policy based on model estimates. Compared to BWS studies, where only one study predicted non-complete choices using model estimates, a larger percentage of DCE studies (42.3%) predicted respondent choices using simulation.

2.4.5 Software

Latent Gold (LG) Choice (26.3%), Sawtooth (21.2%), and NLOGIT (20.2%) were the most common software for segmentation, followed by Stata (14.1%). Sawtooth has been used in the past two decades since segmentation models were first applied to health, while LG Choice, NLOGIT, and Stata started to gain popularity since 2010. Large variations in the choice of test statistics and model specification exist between different software. When testing preference heterogeneity between different classes, Sawtooth users tend to use statistical methods such as ANCOVA, MANCOVA, or *t*-test; LG

Choice users only used Wald test; and no statistical tests was done among Stata users. This is likely due to convenience. For example, Wald test statistics are always reported in the output in LG Choice, which is not the case in other software. In terms of examining respondent characteristics in each class, individual characteristics were included in the segmentation model in 53.8% of studies in Stata, 57.7% of studies in LG Choice, and 81.0% of studies in NLOGIT, but none of the studies in Sawtooth. Simulations were more common with Sawtooth (61.9%) and LG Choice (34.6%) possibly because a more user-friendly option for simulation is provided.

2.5 Study Quality and Implication

2.5.1 PREFS Scores

PREFS scores of the 99 studies ranged from one to five. A majority of the studies scored three (56.6%) or four (28.3%). Only one study scored a one and four studies scored a five. The overall PREFS scores increased overtime by 0.07 per year in the past two decades ($p < 0.001$). Average PREFS scores for studies in the UK and other European countries were significantly higher than the average in the US by 0.72 ($p = 0.011$) and 0.58 ($p = 0.019$), respectively, while other regions were similar to the US ($p > 0.05$).

Only one study did not clearly state the research purpose in relation to preferences. Most of the studies (86.9%) failed to demonstrate that responders were similar to non-responders. Most of the studies (90.9%) clearly explained the preference elicitation methods by including the actual preference question presented to the respondents or referencing it elsewhere. Three-fourths of the studies (74.7%) excluded respondents from the analysis and failed to show that the excluded did not significantly differ from the included. Only four studies (4.0%) did not use significance tests to assess preference results.

2.5.2 Implications

One-third of studies (36.4%) discussed the policy implications of segmentation results, while 43.4% used the results to inform clinical practice. Among them, nine articles provided both policy and clinical recommendations. Preference results were used to direct the design of effective prevention programs in 10.1% of studies. Nine studies provided methodological recommendations. LCA has been applied in clinical studies to understand patient preference heterogeneity and inform clinical practice over the past two decades, whereas the application of LCA in policy decision-making started in 2010 and has been increasing thereafter. No difference in stated-preference methods or the number of classes was found between clinical studies and policy studies ($p > 0.05$).

2.6 Discussion

This study overviews segmentation methods used in preference studies in health. We found that the application of LCA in health-related stated-preference studies has increased dramatically over the past decade in a widening range of countries. Preferences were assessed in a wide variety of topics, including chronic diseases, preventative health behaviors, and social behaviors. Preferences among patients and the general population were the focus of the majority of the studies, followed by the preferences of informal caregivers and healthcare providers. Several studies compared preference across different populations. A large proportion of the studies used segmentation results to support the development of tailored treatments, prevention programs, or direct personalized clinical care to meet various needs in the past two decades. Since 2010, LCA has been increasingly used to inform policy.

2.6.1 Current Practice

As the application of LCA in health is still relatively new, there is little guidance or consensus on how to perform LCA. We summarized six common steps to conduct LCA after reviewing the articles. We found large variations in each step among the studies in our review (Fig. 2-7). In this section, we summarize and compare the practice in current literature to the evidence to date, if available, to identify knowledge gaps and provide guide for researchers who are interested in using segmentation methods to study preference heterogeneity in the health context.

[FIGURE 2-7 INSERT HERE]

Step 1: Choose segmentation methods. The choice of segmentation method depends on the preference elicitation method used in the study. While all rating studies segmented respondents with distance-minimizing procedures such as k-mean cluster analysis, clues, and pamk, choice-based conjoint analyses (e.g., DCE and BWS) predominantly used LCL models. Among the latter, only 10% of the studies adjusted for scales, although failing to account for heterogeneity in variance (i.e. scales) in these LCL models has been shown to lead to biased estimates and predictions (Magidson and Vermunt, 2007).

Step 2: Select optimal number of classes. LCA involves comparing segmentation models with various pre-specified numbers of classes to identify the appropriate number of classes based on model fit. Common criteria for model selection in k-mean cluster analyses include Ward's method and the maximized Kappa statistic. Researchers can also consider class interpretability and use respondents' self-reported values for individual attributes to validate the results.

A wide range of information criteria, as well as statistics such as LL, Entropy R-squared, and McFadden's R-squared, have been used to select the best-fitted LCL models. Although BIC has been shown to have better performance than other information criteria such as AIC (Nylund et al., 2007), only half of the studies used it for model selection. As the information criteria and other statistics only evaluate statistical performance of the model, a few studies adjusted the segmentation results based on class size, interpretability, and variation between classes to ensure policy or clinical relevance.

Looking across the studies, the number of classes identified was not affected by preference elicitation methods, the number of attributes and levels included in the survey, segmentation methods, or model selection criteria. However, larger sample size leads to more classes. This could be because increasing the number of respondents introduces greater preference heterogeneity, or because larger sample sizes have more statistical power to detect preference variation. Given that half of the studies had fewer than 300 respondents, sample size could be the reason why 80% of studies in this review reported two or three classes and that we did not find any differences in the number of preference types between different populations (e.g., patients, caregivers) and age groups.

Step 3: Compare LCA to alternative models. LCA assumes that preferences are distributed discretely into like clusters (segments) (McFadden, 1986; Swait, 1994), which may not always be true. Researchers should therefore check this assumption by testing the existence of distinct clusters in the data and comparing LC model to alternative model specifications to detect the underlying distribution of preferences. Two-thirds of studies in this review included alternative models in their analyses, but only half (or 7% among BWS) statistically compared the performance of different models. Moreover, while many studies compared the segmentation model to the aggregated model to check the existence of distinct classes, only a small number of studies tested for alternative preference distribution assumptions. Comparison between alternative models should be encouraged because misspecification of preference distribution may lead to biased results. However, an increasing use of Hierarchical Bayes to estimate individual preference weights reduces reliance on distribution assumptions.

Step 4: Compare preferences between classes. If the discrete-distribution assumption holds, preference heterogeneity between classes should be tested statistically because variation in the magnitude of preference estimates may not be significant, depending on class size and estimate standard errors. In addition, a test of parameters between classes may also shed light on the distribution of preference estimates. Only 30% of studies in this review compared preference between classes statistically, possibly because studies with small sample sizes were underpowered to detect heterogeneity between classes (i.e., there were large standard errors in some classes). It could also be influence by the type of software used for analyses. For example, Wald tests for equal preference estimates across classes were performed more by LG Choice users where these estimates are included in the output of LCA. In comparison, statistical tests across classes are more difficult to perform in Stata, and none of the Stata users statistically tested cross-class preference heterogeneity. Packages should be developed to allow easy access to post-estimation tests in all software, and researchers should choose software more strategically for LCAs.

Step 5: Identify class characteristics. Researchers and policymakers are often interested in understanding class membership in order to develop interventions and policies that can be tailored to these groups. Including individual covariates (i.e., socio demographic and other relevant variables) in the segmentation model to help predict class membership has been shown to improve model identification, fit, and interpretation (Boxall and Adamowicz, 2002; Huang and Bandeen-Roche, 2004; Lanza and Rhoades, 2011; Yang and Yang, 2007), and is therefore recommended. About half of the studies in the literature included covariates in the model, and some tested the impact of including covariates on model performance. Covariate selection was often done using backward selection method or based on theory. Sawtooth users should be particularly encouraged to include covariates in the segmentation model given that none of the Sawtooth studies considered covariates.

Among studies that did not include covariates in the segmentation model, a majority conducted post-estimation analyses to examine class-specific characteristics using chi-square tests, ANOVA, MANOVA, or logistic regression. Deal (2014) showed a more systematic approach (i.e., random forest) to identify individual variables associated with class membership after model estimation, which may provide a useful guide for researchers.

Step 6: Predict choices or market shares. LCL models estimate class-specific preference weights that determine the probabilities that each class will choose a given treatment or service. When individual covariates are included, LCL models also estimate the class membership probability for an individual. Researchers can therefore predict individual choices based on participant characteristics, or estimate market uptake based on market socio demographic composition. Prediction allows further cost-benefit analyses that compare the value of the treatment or service to the cost of development given market uptake to support decision-making.

Predicting individual choices is challenging when individual covariates are not included in the model, but market shares can be estimated based on class-specific preference weights and segment size or using simulation. Randomized first-choice simulation is the most common method used for prediction, especially among Sawtooth users. While one-third of studies in the literature have performed prediction, few used the results for cost-benefit analyses. Future research should apply segmentation results to policy or program development.

The quality of LCA studies improved over time. A majority of the studies clearly stated the study objective in relation to preferences and described preference elicitation methods and procedures. However, many studies failed to explain the details of statistical analysis such as how the model was selected when multiple criteria led to different conclusions (as is the case in most LCAs), how class characteristics were compared, and the statistics used to test for group differences. Many studies did not justify the choice of sample size, model specification, model selection criteria, and covariates to include in model. In terms of generalizability, a majority of the studies did not provide sufficient evidence to show that respondents in the analysis were representative of the study population. As the application of LCA in health expands, researchers should continue improving the quality as well as the reporting of LCA studies.

2.6.2 Limitations

There are two major limitations in this review. First, it is challenging to determine the boundary of public health relevance. Many topics, such as the choice of organic or genetically modified foods, were excluded due to undetermined health impacts. Future research could further examine LCAs in these fields, but the segmentation methods used in these studies are likely similar. Second, there could be publication bias in the literature. In our review, only one study stated that LCA was conducted but not reported because the aggregated model had better fit. It is possible that studies not included in this review performed segmentation but only reported results from other models (e.g., CL, MXL) due to better fit. Future research should be more transparent about methods used in the study.

2.7 Conclusion

LCA has been increasingly used to study preference heterogeneity in health. It has been applied in a wide range of medical conditions and health areas to support the development of treatments or prevention programs. However, there is little consensus and guidance on appropriate application. With the increasing emphasis on patient preference from regulatory agencies and patient-centered care in clinical settings, there will be higher demand to study patient preference and preference heterogeneity. More guidance is needed to improve the quality of LCA studies in health to meet this demand and support policy development and clinical practice.

Reference

Adamowicz W, Swait J, Boxall P, et al. Perceptions versus objective measures of environmental quality in combined revealed and stated preference models of environmental valuation. Journal of Environmental Economics and Management. 1997;32(1):65-84.

Bateman IJ, Carson RT, Day B, et al. Economic valuation with stated preference techniques: a manual. Cheltenham: Edward Elgar Publishing Ltd; 2002.

Boxall PC, Englin J, and Adamowicz WL. Valuing Aboriginal artifacts: A combined revealed-stated preference approach. Journal of Environmental Economics and Management. 2003;45(2):213-30.

Boxall PC and Adamowicz WL. Understanding heterogeneous preferences in random utility models: a latent class approach. Environmental and Resource Economics. 2002;23:421–446.

Califf RM. Benefit-risk assessments at the US Food and Drug Administration: finding the balance. JAMA. 2017. doi: 10.1001/jama.2017.0410. [Epub ahead of print]

Center for Devices and Radiological Health (CDRH) and Center for Biologics Evaluation and Research (CBER). Patient preference information – submission, review in PMAs, HDE applications, and *de novo* requests, and inclusion in device labeling: draft guidance for industry, Food and Drug Administration staff, and other stakeholders. Issued on May 18, 2015.

Clark MD, Determann D, Petrou S, et al. Discrete choice experiments in health economics: a review of the literature. Pharmacoeconomics. 2014;32(9):883-902.

de Bekker-Grob EW, Ryan M, and Gerard K. Discrete choice experiments in health economics: a review of the literature. Health Economics. 2012;21(2):145-72.

Deal K. Segmenting patients and physicians using preferences from discrete choice experiments. Patient. 2014;7:5-21.

Hilger J and Hanemann M. Heterogeneous preferences for water quality: A finite mixture model of beach recreation in Southern California. San Diego: Research Final Reports, California Sea Grant College Program, UC San Diego; 2006.

Huang GH, and Bandeen-Roche K. Building an identifiable latent class model with covariate effects on underlying and measured variables. Psychometrika. 2004;69:5–32.

Iraguen P, and de Dios Ortuzar J. Willingness-to-pay for reducing fatal accident risk in urban areas: an internet-based web page stated preference survey. Accident Analysis and Prevention. 2004;36(4):513-24.

Lanza ST, and Rhoades BL. Latent class analysis: An alternative perspective on subgroup analysis in prevention and treatment. Prevention Science. 2011;1–12.

Magidson J, and Vermunt JK. Removing the scale factor confound in multinomial logit choice models to obtain better estimates of preference. Sawtooth Softw Conf 2007, 1–18.

McFadden D. The choice theory approach to market research. Marketing Science. 1986;5(4):275-97.

Morey E and Greer Rossmann K. Using stated-preference questions to investigate variations in willingness to pay for preserving marble monuments: Classic heterogeneity, random parameters, and mixture models. Journal of Cultural Economics. 2003;27(3):215-29.

Nylund KL, Asparouhov T, and Muthén BO. Deciding on the number of classes in latent class analysis and growth mixture modeling: a Monte Carlo simulation study. Structural Equation Modeling. 2007;14:535–569.

Swait J. A structural equation model of latent segmentation and product choice for cross-sectional revealed preference choice data. Journal of Retailing and Consumer Services. 1994;1(2):77-89.

Vass C, Rigby D, Payne K. The role of research methods in discrete choice experiments: a systematic review and survey of authors. Medical Decision Making. 2017. doi: 10.1177/0272989X16683934. [Epub ahead of print]

Yang CC, and Yang CC. Separating latent classes by information criteria. Journal of Classification. 2007;24(2):183–203.

Figure 2-1 PRISMA Diagram

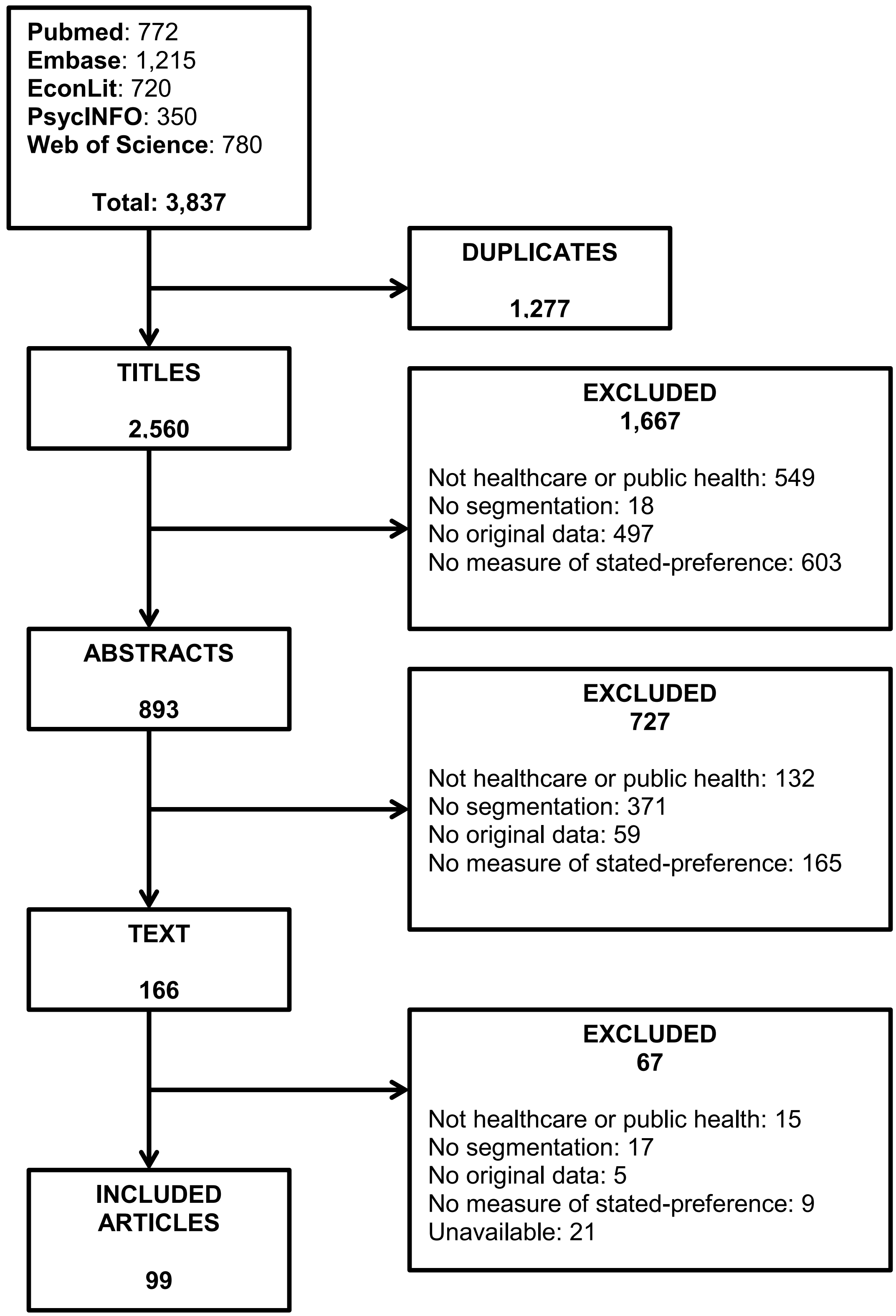
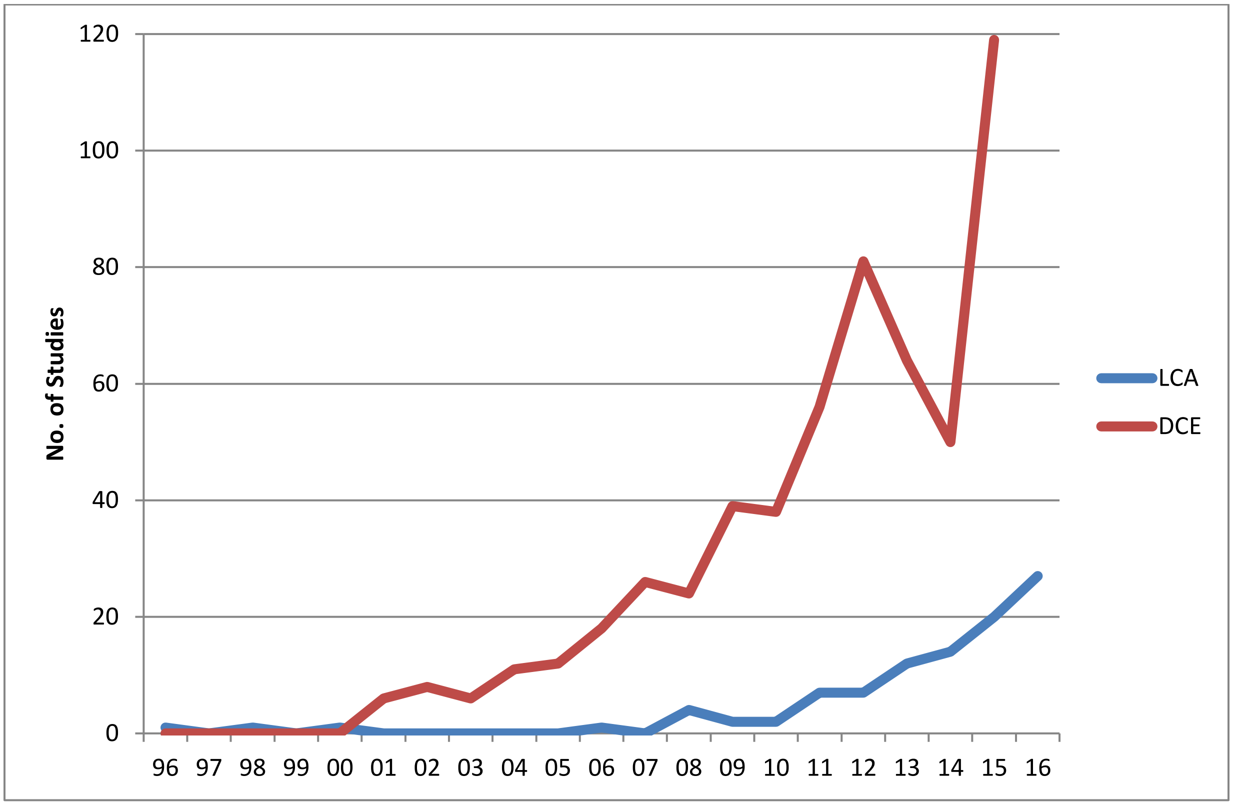


Figure 2-2 Number of segmentation studies and discrete-choice studies by year



* The number of DCE per year is cited from Vass et al. (2017).

Figure 2-3 Number of attributes included in the studies by stated-preference methods

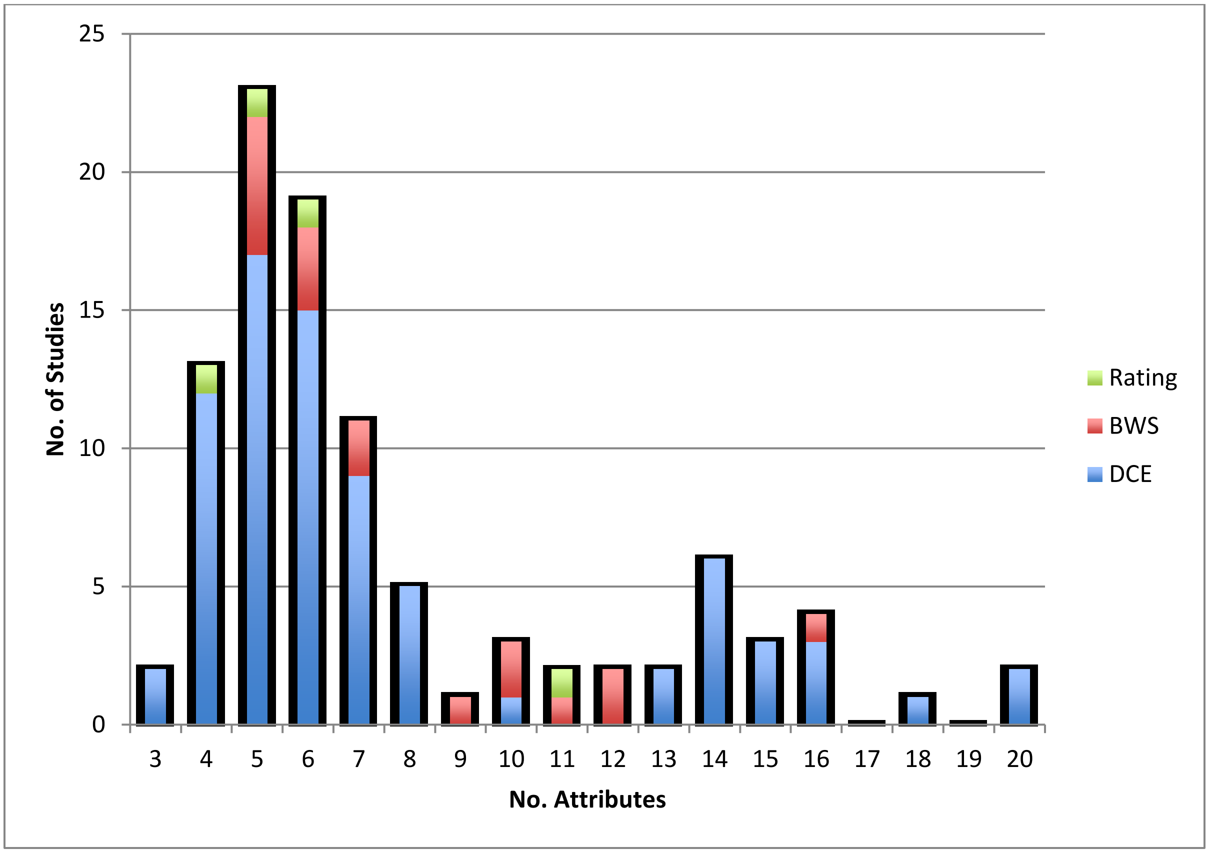


Figure 2-4 Distribution of sample size in latent class analysis in the studies

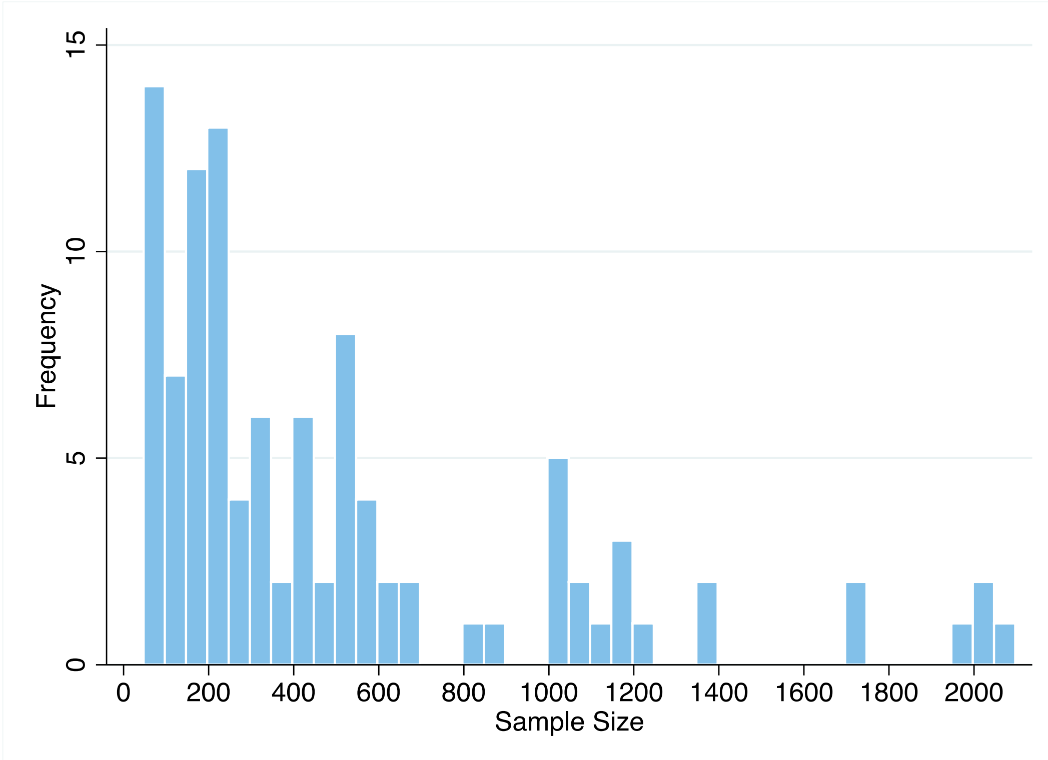


Figure 2-5 Scatter plot between sample size and number of classes reported in studies

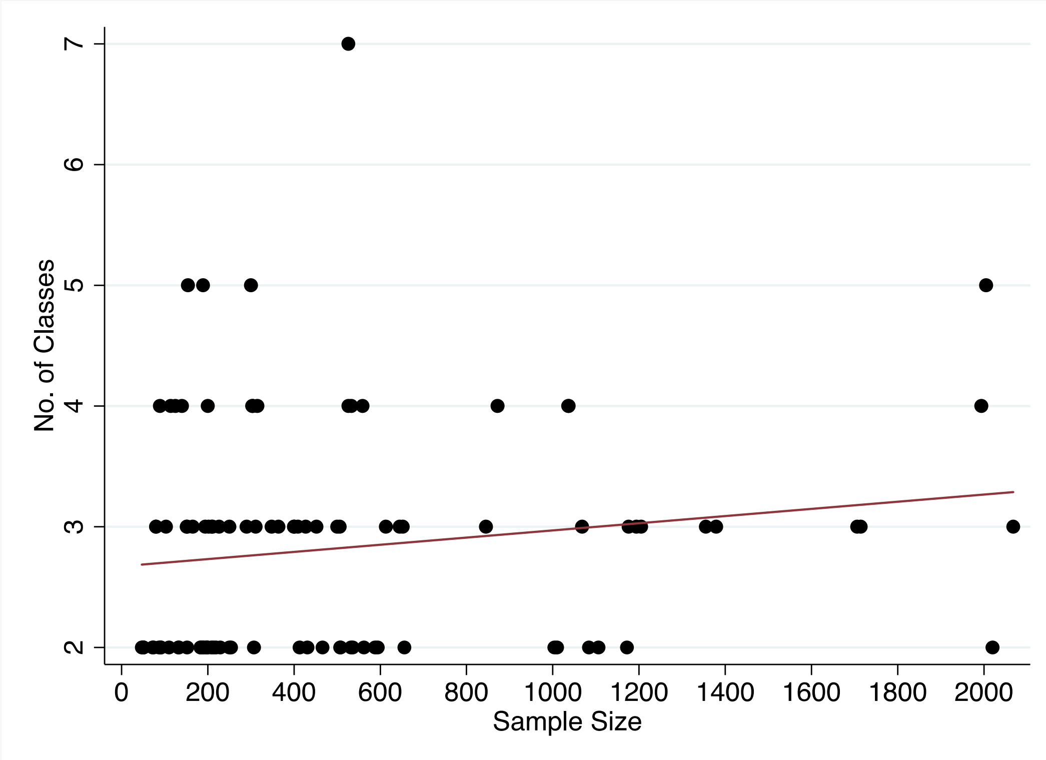
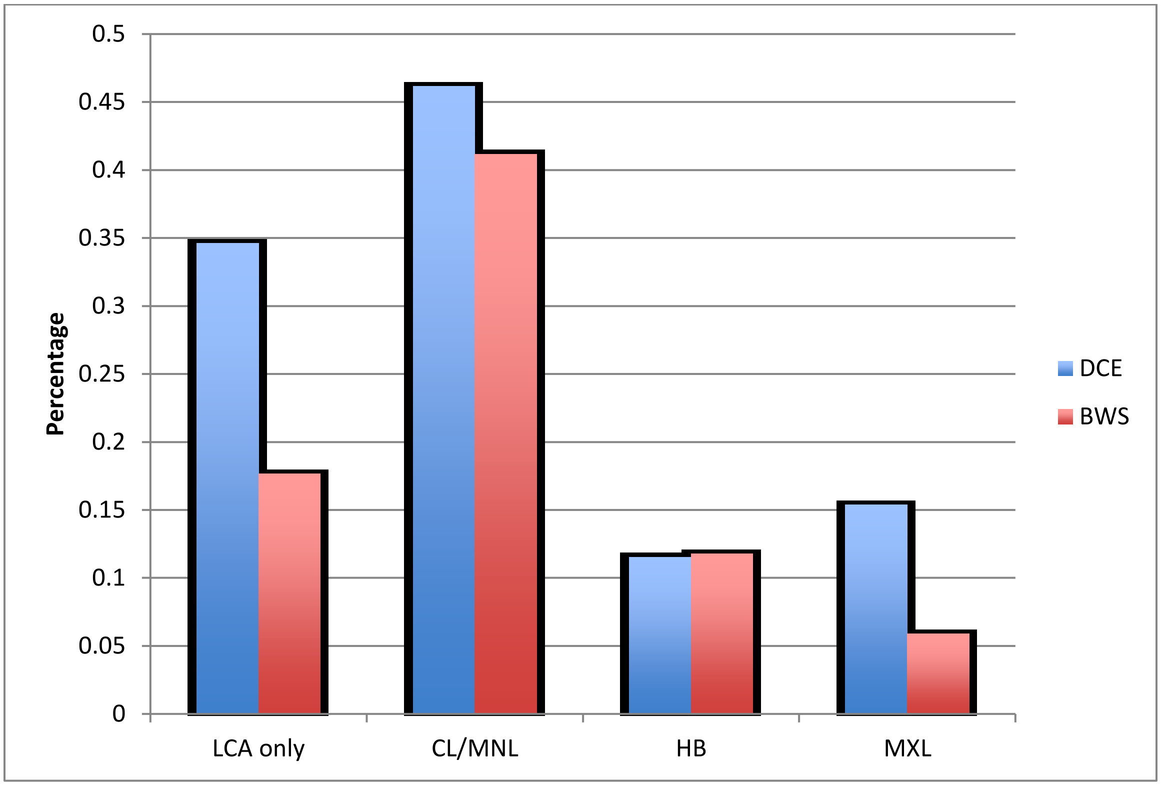


Figure 2-6 Alternative models used in the studies in addition to segmentation model



LCA: latent class analysis; CL: conditional logit; MNL: multinomial logit; HB: hierarchical bayes; MXL: mixed logit; DCE: discrete choice experiment; BWS: best-worst scaling.

Figure 2-7 Six steps for segmentation analysis with preference data

Choose Segmentation Methods

- Distance-minimizing procedures (e.g. k-mean, clues, pamk)
- Probabilistic models (e.g. LCL)

Select Optimal Number of Classes

- K-mean: Ward's method, maximized Kappa, external validation
- LCL: information criteria (e.g. BIC), LL, Entropy R^2, McFadden's adjusted R^2, class size, interpretability, holdout task hit rate

Compare LCA to Alternative Models

- Alternative models: aggregate model, MXL, HB
- Criteria: information criteria (e.g. BIC), LL, McFadden's adjusted R^2, prediction hit rate

Compare Preference between Classes

- K-mean: ANCOVA
- LCL: Wald test, ANCOVA, MANCOVA, Dunnett's C test, Tukey comparison, t-test

Identify Class Characteristics

- Include individual characteristics in segmentation model
- Post-estimation analyses: random forest

Predict Choices or Market Shares

- Scenario analyses
- Randomized first-choice simulation

Table 2-1 Linear regression of the number of attributes and levels on study characteristics

	No. of Attributes		No. of Levels	
	Coef.	(SE)	Coef.	(SE)
No. of attributes	-	-	-0.03	(0.04)
Country of origin				
Australia	0.19	(1.36)	0.19	(0.48)
Canada	3.15	(1.37)*	1.33	(0.40)**
UK	-1.20	(1.15)	0.94	(0.48)*
Other EU	-0.95	(1.15)	0.42	(0.36)
Other	-1.53	(1.08)	0.35	(0.49)
US	*Reference*			
Study population				
Patient	-2.55	(1.01)*	0.08	(0.32)
Clinician	-0.60	(1.13)	-0.07	(0.37)
Caregiver	2.05	(1.21)	-0.25	(0.24)
General population	-3.29	(1.12)**	0.78	(0.50)
Age group				
≤18 years	0.38	(1.12)	0.09	(0.44)
>18 years	0.79	(0.75)	-0.03	(0.36)
>65 years	0.86	(0.96)	-0.28	(0.45)
No age restriction	*Reference*			
SP methods				
Rating	1.03	(0.87)	-0.37	(0.78)
BWS	-0.07	(1.13)	-0.65	(0.30)*
DCE	*Reference*			

* p<0.05; ** p<0.01; *** p<0.001

a. Huber-White standard errors were used to adjust for heteroscedasticity.

Table 2-2 Linear regression of sample size on study characteristics

	Coef.	(SE)
No. of attributes	58.23	(11.65)***
No. of levels	53.40	(39.02)
Country of origin		
Australia	547.56	(218.07)*
Canada	195.36	(145.92)
UK	218.95	(174.60)
Other EU	259.63	(132.14)
Other	75.10	(127.70)
US	*Reference*	
Study population		
Patient	-46.36	(98.68)
Clinician	-29.91	(115.37)
Caregiver	49.71	(102.43)
General population	360.12	(144.26)*
Age group		
≤18 years	242.49	(193.14)
>18 years	140.77	(111.11)
>65 years	-207.26	(137.57)
No age restriction	*Reference*	
SP methods		
Rating	-133.57	(189.33)
BWS	-162.67	(92.72)
DCE	*Reference*	

* $p<0.05$; ** $p<0.01$; *** $p<0.001$

a. Huber-White standard errors were used to adjust for heteroscedasticity.

Table 2-3 Linear regression of the number of classes on study characteristics

	Coef.	(SE)
Sample size (in 100s)	0.05	(0.02)*
No. of attributes	-0.01	(0.03)
No. of levels	0.04	(0.08)
Country of origin		
Australia	0.89	(0.63)
Canada	-0.21	(0.35)
UK	-0.05	(0.38)
Other EU	0.23	(0.44)
Other	-0.05	(0.35)
US	*Reference*	
Study population		
Patient	-0.25	(0.29)
Clinician	0.25	(0.34)
Caregiver	0.17	(0.29)
General population	-0.64	(0.37)
Age group		
≤18 years	-0.51	(0.37)
>18 years	0.12	(0.27)
>65 years	0.72	(0.38)
No age restriction	*Reference*	
SP methods		
Rating	-0.45	(0.89)
BWS	-0.13	(0.36)
DCE	*Reference*	
Segmentation methods		
LCL	*Reference*	
Scale-adjusted LCL	0.40	(0.51)
K-mean	0.59	(0.85)
Model selection criteria		
AIC	-0.06	(0.28)
BIC	0.22	(0.20)
Log-likelihood	0.45	(0.30)
Interpretability	0.29	(0.25)

* $p<0.05$; ** $p<0.01$; *** $p<0.001$

a. Huber-White standard errors were used to adjust for heteroscedasticity.

Appendix A. Included articles in the systematic review

1. Ammi M, Peyron C. Heterogeneity in general practitioners' preferences for quality improvement programs: a choice experiment and policy simulation in France. Health Economics Review. 2016;6(1):44.
2. Aspinall PA, Johnson ZK, Azuara-Blanco A, Montarzino A, Brice R, Vickers A. Evaluation of quality of life and priorities of patients with glaucoma. Investigative ophthalmology & visual science. 2008;49(5):1907-15.
3. Bailey K, Cunningham C, Pemberton J, Rimas H, Morrison KM. Understanding Academic Clinicians' Decision Making for the Treatment of Childhood Obesity. Childhood Obesity. 2015;11(6):696-706.
4. Becker MP, Christensen BK, Cunningham CE, Furimsky I, Rimas H, Wilson F et al. Preferences for early intervention mental health services: A discrete-choice conjoint experiment. Psychiatric Services. 2016;67(2):184-91.
5. Brown DS, Poulos C, Johnson FR, Chamiec-Case L, Messonnier ML. Adolescent girls' preferences for HPV vaccines: a discrete choice experiment. Preference Measurement in Health. Emerald Group Publishing Limited; 2014. p. 93-121.
6. Brown ZS, Kramer RA, Ocan D, Oryema C. Household perceptions and subjective valuations of indoor residual spraying programmes to control malaria in northern Uganda. Infectious Diseases of Poverty. 2016;5(1):100.
7. Cai Q, Wan F, Dong X, Liao X, Zheng J, Wang R et al. Fertility clinicians and infertile patients in China have different preferences in fertility care. Human Reproduction. 2014;29(4):712-9.
8. Carroll FE, Al‐Janabi H, Flynn T, Montgomery AA. Women and their partners' preferences for Down's syndrome screening tests: a discrete choice experiment. Prenatal diagnosis. 2013;33(5):449-56.
9. Cunningham CE, Barwick M, Short K, Chen Y, Rimas H, Ratcliffe J et al. Modeling the mental health practice change preferences of educators: A discrete-choice conjoint experiment. School mental health. 2014;6(1):1-14.
10. Cunningham CE, Bruce BS, Snowdon AW, Chen Y, Kolga C, Piotrowski C et al. Modeling improvements in booster seat use: a discrete choice conjoint experiment. Accident Analysis & Prevention. 2011;43(6):1999-2009.
11. Cunningham CE, Chen Y, Deal K, Rimas H, McGrath P, Reid G et al. The interim service preferences of parents waiting for children's mental health treatment: A discrete choice conjoint experiment. Journal of abnormal child psychology. 2013;41(6):865-77.
12. Cunningham CE, Deal K, Neville A, Rimas H, Lohfeld L. Modeling the problem-based learning preferences of McMaster University undergraduate medical students using a discrete choice conjoint experiment. Advances in Health Sciences Education. 2006;11(3):245-66.
13. Cunningham CE, Deal K, Rimas H, Buchanan DH, Gold M, Sdao-Jarvie K et al. Modeling the information preferences of parents of children with mental health problems: A discrete choice conjoint experiment. Journal of abnormal child psychology. 2008;36(7):1123.
14. Cunningham CE, Deal K, Rimas H, Campbell H, Russell A, Henderson J et al. Using conjoint analysis to model the preferences of different patient segments for attributes of patient-centered care. The Patient: Patient-Centered Outcomes Research. 2008;1(4):317-

30.
15. Cunningham CE, Deal K, Rimas H, Chen Y, Buchanan DH, Sdao-Jarvie K. Providing information to parents of children with mental health problems: a discrete choice conjoint analysis of professional preferences. Journal of abnormal child psychology. 2009;37(8):1089.
16. Cunningham CE, Henderson J, Niccols A, Dobbins M, Sword W, Chen Y et al. Preferences for evidence‐based practice dissemination in addiction agencies serving women: a discrete‐choice conjoint experiment. Addiction. 2012;107(8):1512-24.
17. Cunningham CE, Hutchings T, Henderson J, Rimas H, Chen Y. Modeling the hospital safety partnership preferences of patients and their families: a discrete choice conjoint experiment. Patient preference and adherence. 2016;10:1359.
18. Cunningham CE, Kostrzewa L, Rimas H, Chen Y, Deal K, Blatz S et al. Modeling Organizational Justice Improvements in a Pediatric Health Service. The Patient-Patient-Centered Outcomes Research. 2013;6(1):45-59.
19. Cunningham CE, Rimas H, Chen Y, Deal K, McGrath P, Lingley-Pottie P et al. Modeling Parenting Programs as an Interim Service for Families Waiting for Children's Mental Health Treatment. Journal of Clinical Child & Adolescent Psychology. 2015;44(4):616-29.
20. Cunningham CE, Vaillancourt T, Cunningham LJ, Chen Y, Ratcliffe J. Modeling the bullying prevention program design recommendations of students from grades five to eight: a discrete choice conjoint experiment. Aggressive behavior. 2011;37(6):521-37.
21. Cunningham CE, Vaillancourt T, Rimas H, Deal K, Cunningham L, Short K et al. Modeling the bullying prevention program preferences of educators: A discrete choice conjoint experiment. Journal of abnormal child psychology. 2009;37(7):929-43.
22. Cunningham CE, Walker JR, Eastwood JD, Westra H, Rimas H, Chen Y et al. Modeling mental health information preferences during the early adult years: a discrete choice conjoint experiment. Journal of health communication. 2014;19(4):413-40.
23. de Bekker-Grob EW, Rose J, Donkers B, Essink-Bot M-L, Bangma C, Steyerberg E. Men's preferences for prostate cancer screening: a discrete choice experiment. British journal of cancer. 2013;108(3):533-41.
24. de-Magistris T, Lopéz-Galán B. Consumers' willingness to pay for nutritional claims fighting the obesity epidemic: the case of reduced-fat and low salt cheese in Spain. Public health. 2016;135:83-90.
25. Deal K. Segmenting patients and physicians using preferences from discrete choice experiments. The Patient-Patient-Centered Outcomes Research. 2014;7(1):5-21.
26. Deal K, Keshavjee K, Troyan S, Kyba R, Holbrook AM. Physician and patient willingness to pay for electronic cardiovascular disease management. International journal of medical informatics. 2014;83(7):517-28.
27. Determann D, Korfage IJ, Fagerlin A, Steyerberg EW, Bliemer MC, Voeten HA et al. Public preferences for vaccination programmes during pandemics caused by pathogens transmitted through respiratory droplets–a discrete choice experiment in four European countries, 2013 Alternative title (right-running): www. eurosurveillance. org–Determann et al. Euro Surveill. 2016;21(22):30247.
28. Determann D, Korfage IJ, Lambooij MS, Bliemer M, Richardus JH, Steyerberg EW et al. Acceptance of vaccinations in pandemic outbreaks: a discrete choice experiment. PloS one. 2014;9(7):e102505.

29. Determann D, Lambooij MS, de Bekker-Grob EW, Hayen AP, Varkevisser M, Schut FT et al. What health plans do people prefer? The trade-off between premium and provider choice. Social Science & Medicine. 2016;165:10-8.
30. Dong D, Ozdemir S, Bee YM, Toh S-A, Bilger M, Finkelstein E. Measuring high-risk patients' preferences for pharmacogenetic testing to reduce severe adverse drug reaction: a discrete choice experiment. Value in Health. 2016;19(6):767-75.
31. Erdem S, Campbell D, Thompson C. Elimination and selection by aspects in health choice experiments: Prioritising health service innovations. Journal of health economics. 2014;38:10-22.
32. Erdem S, Thompson C. Prioritising health service innovation investments using public preferences: a discrete choice experiment. BMC health services research. 2014;14(1):360.
33. Feudtner C, Walter JK, Faerber JA, Hill DL, Carroll KW, Mollen CJ et al. Good-parent beliefs of parents of seriously ill children. JAMA pediatrics. 2015;169(1):39-47.
34. Finkelstein E, Malhotra C, Chay J, Ozdemir S, Chopra A, Kanesvaran R. Impact of Treatment Subsidies and Cash Payouts on Treatment Choices at the End of Life. Value in Health. 2016;19(6):788-94.
35. Flynn TN, Bilger M, Malhotra C, Finkelstein EA. Are Efficient Designs Used in Discrete Choice Experiments Too Difficult for Some Respondents? A Case Study Eliciting Preferences for End-of-Life Care. PharmacoEconomics. 2016;34(3):273-84.
36. Flynn TN, Huynh E, Peters TJ, Al‐Janabi H, Clemens S, Moody A et al. Scoring the Icecap‐a Capability Instrument. Estimation of a UK General Population Tariff. Health Economics. 2015;24(3):258-69.
37. Flynn TN, Louviere JJ, Peters TJ, Coast J. Using discrete choice experiments to understand preferences for quality of life. Variance-scale heterogeneity matters. Social science & medicine. 2010;70(12):1957-65.
38. Flynn TN, Peters TJ, Coast J. Quantifying response shift or adaptation effects in quality of life by synthesising best-worst scaling and discrete choice data. Journal of choice modelling. 2013;6:34-43.
39. Fraenkel L, Lim J, Garcia-Tsao G, Reyna V, Monto A. Examining hepatitis C virus treatment preference heterogeneity using segmentation analysis: treat now or defer? Journal of clinical gastroenterology. 2016;50(3):252-7.
40. Fraenkel L, Lim J, Garcia-Tsao G, Reyna V, Monto A, Bridges JF. Variation in treatment priorities for chronic hepatitis C: a latent class analysis. The Patient-Patient-Centered Outcomes Research. 2016;9(3):241-9.
41. Fraenkel L, Suter L, Cunningham CE, Hawker G. Understanding Preferences for Disease‐Modifying Drugs in Osteoarthritis. Arthritis care & research. 2014;66(8):1186-92.
42. Fuller-Tyszkiewicz M, Reynard K, Skouteris H, McCabe M. An examination of the contextual determinants of self-objectification. Psychology of Women Quarterly. 2012;36(1):76-87.
43. Gallego G, Dew A, Lincoln M, Bundy A, Chedid RJ, Bulkeley K et al. Should I stay or should I go? Exploring the job preferences of allied health professionals working with people with disability in rural Australia. Human resources for health. 2015;13(1):53.
44. Gong J, Zhang Y, Feng J, Zhang W, Yin W, Wu X et al. How best to obtain consent to thrombolysis Individualized decision-making. Neurology. 2016;86(11):1045-52.

45. Goossens LM, Utens CM, Smeenk FW, Donkers B, van Schayck OC, Rutten-van Mölken MP. Should I stay or should I go home? A latent class analysis of a discrete choice experiment on hospital-at-home. Value in Health. 2014;17(5):588-96.
46. Grindrod KA, Marra CA, Colley L, Tsuyuki RT, Lynd LD. Pharmacists' preferences for providing patient-centered services: a discrete choice experiment to guide health policy. Annals of Pharmacotherapy. 2010;44(10):1554-64.
47. Grisolía JM, Longo A, Hutchinson G, Kee F. Applying Health Locus of Control and Latent Class Modelling to food and physical activity choices affecting CVD risk. Social Science & Medicine. 2015;132:1-10.
48. Guo N, Marra CA, FitzGerald JM, Elwood RK, Anis AH, Marra F. Patient preference for latent tuberculosis infection preventive treatment: a discrete choice experiment. Value in Health. 2011;14(6):937-43.
49. Haegeli P, Gunn M, Haider W. Identifying a high-risk cohort in a complex and dynamic risk environment: Out-of-bounds skiing—An example from avalanche safety. Prevention Science. 2012;13(6):562-73.
50. Hazlewood GS, Bombardier C, Tomlinson G, Thorne C, Bykerk VP, Thompson A et al. Treatment preferences of patients with early rheumatoid arthritis: a discrete-choice experiment. Rheumatology. 2016;55(11):1959-68.
51. Hifinger M, Hiligsmann M, Ramiro S, Watson V, Severens J, Fautrel B et al. Economic considerations and patients' preferences affect treatment selection for patients with rheumatoid arthritis: a discrete choice experiment among European rheumatologists. Annals of the rheumatic diseases. 2016:annrheumdis-2016-209202.
52. Hofman R, de Bekker-Grob EW, Richardus JH, de Koning HJ, van Ballegooijen M, Korfage IJ. Have preferences of girls changed almost 3 years after the much debated start of the HPV vaccination program in the Netherlands? A discrete choice experiment. PloS one. 2014;9(8):e104772.
53. Hole AR. Modelling heterogeneity in patients’ preferences for the attributes of a general practitioner appointment. Journal of health economics. 2008;27(4):1078-94.
54. Hole AR, Norman R, Viney R. Response patterns in health state valuation using endogenous attribute attendance and latent class analysis. Health economics. 2014.
55. Howard K, Jan S, Rose JM, Wong G, Craig JC, Irving M et al. Preferences for Policy Options for Deceased Organ Donation for Transplantation: A Discrete Choice Experiment. Transplantation. 2016;100(5):1136-48.
56. Jarvis W, Pettigrew S. The relative influence of alcohol warning statement type on young drinkers’ stated choices. Food Quality and Preference. 2013;28(1):244-52.
57. Kan HJ, de Bekker-Grob EW, Van Marion ES, Van Oijen GW, van Nieuwenhoven CA, Zhou C et al. Patients’ preferences for treatment for Dupuytren’s disease: A discrete choice experiment. Plastic and reconstructive surgery. 2016;137(1):165-73.
58. Kuzmanovic M, Vujosevic M, Martic M. Using conjoint analysis to elicit patients’ preferences for public primary care service in Serbia. HealthMED. 2012;6(2):496-504.
59. Lacanilao RD, Cash SB, Adamowicz WL. Heterogeneous consumer responses to snack food taxes and warning labels. Journal of Consumer Affairs. 2011;45(1):108-22.
60. Lagarde M. Investigating attribute non‐attendance and its consequences in choice experiments with latent class models. Health economics. 2013;22(5):554-67.
61. Lagarde M, Erens B, Mays N. Determinants of the choice of GP practice registration in England: evidence from a discrete choice experiment. Health Policy. 2015;119(4):427-

36.
62. Lau MA, Colley L, Willett BR, Lynd LD. Employee's preferences for access to mindfulness-based cognitive therapy to reduce the risk of depressive relapse—A discrete choice experiment. Mindfulness. 2012;3(4):318-26.
63. Liang D, Tang C-X. The specialty choice of medical students in China: a stated preference experiment. BMC medical education. 2016;16(1):107.
64. Lock J, Bekker‐Grob E, Urhan G, Peters M, Meijer K, Brons P et al. Facilitating the implementation of pharmacokinetic‐guided dosing of prophylaxis in haemophilia care by discrete choice experiment. Haemophilia. 2016;22(1):e1-e10.
65. Lynd LD, Traboulsee A, Marra CA, Mittmann N, Evans C, Li KH et al. Quantitative analysis of multiple sclerosis patients' preferences for drug treatment: a best–worst scaling study. Therapeutic Advances in Neurological Disorders. 2016;9(4):287-96.
66. Mandeville KL, Ulaya G, Lagarde M, Muula AS, Dzowela T, Hanson K. The use of specialty training to retain doctors in Malawi: A discrete choice experiment. Social Science & Medicine. 2016;169:109-18.
67. Mentzakis E, Ryan M, McNamee P. Using discrete choice experiments to value informal care tasks: exploring preference heterogeneity. Health economics. 2011;20(8):930-44.
68. Mentzakis E, Stefanowska P, Hurley J. A discrete choice experiment investigating preferences for funding drugs used to treat orphan diseases: an exploratory study. Health Economics, Policy and Law. 2011;6(03):405-33.
69. Miners A, Llewellyn C, Cooper V, Youssef E, Pollard A, Lagarde M et al. A discrete choice experiment to assess people living with HIV's (PLWHIV's) preferences for GP or HIV clinic appointments. Sex Transm Infect. 2017;93(2):105-11.
70. Mühlbacher AC, Bethge S. Patients' preferences: a discrete-choice experiment for treatment of non-small-cell lung cancer. The European Journal of Health Economics. 2015;16(6):657-70.
71. Naik-Panvelkar P, Armour C, Rose JM, Saini B. Patient preferences for community pharmacy asthma services. Pharmacoeconomics. 2012;30(10):961-76.
72. Najafzadeh M, Gagne JJ, Choudhry NK, Polinski JM, Avorn J, Schneeweiss SS. Patients' Preferences in Anticoagulant Therapy. Circulation: Cardiovascular Quality and Outcomes. 2014;7(6):912-9.
73. Najafzadeh M, Lynd LD, Davis JC, Bryan S, Anis A, Marra M et al. Barriers to integrating personalized medicine into clinical practice: a best-worst scaling choice experiment. Genetics in medicine. 2012;14(5):520-6.
74. Ng X, Bridges JF, Ross MM, Frosch E, Reeves G, Cunningham CE. A Latent Class Analysis to Identify Variation in Caregivers' Preferences for their Child's Attention-Deficit/Hyperactivity Disorder Treatment: Do Stated Preferences Match Current Treatment? The Patient-Patient-Centered Outcomes Research. 2016:1-12.
75. O'Hara NN, Roy L, O'Hara LM, Spiegel JM, Lynd LD, FitzGerald JM et al. Healthcare worker preferences for active tuberculosis case finding programs in South Africa: A best-worst scaling choice experiment. PloS one. 2015;10(7):e0133304.
76. Paolucci F, Mentzakis E, Defechereux T, Niessen LW. Equity and efficiency preferences of health policy makers in China—a stated preference analysis. Health policy and planning. 2014:czu123.
77. Pfarr C, Schmid A. Redistribution through social health insurance: evidence on

citizen preferences. The European Journal of Health Economics. 2016;17(5):611-28.
78. Ratcliffe J, Huynh E, Chen G, Stevens K, Swait J, Brazier J et al. Valuing the Child Health Utility 9D: Using profile case best worst scaling methods to develop a new adolescent specific scoring algorithm. Social Science & Medicine. 2016;157:48-59.
79. Rischatsch M. Who joins the network? Physicians' resistance to take budgetary co-responsibility. Journal of health economics. 2015;40:109-21.
80. Rosato R, Testa S, Oggero A, Molinengo G, Bertolotto A. Quality of life and patient preferences: identification of subgroups of multiple sclerosis patients. Quality of Life Research. 2015;24(9):2173-82.
81. Singh J, Cuttler L, Shin M, Silvers J, Neuhauser D. Medical decision-making and the patient: understanding preference patterns for growth hormone therapy using conjoint analysis. Medical Care. 1998;36(8):AS31-AS45.
82. Skedgel C, Wailoo A, Akehurst R. Societal preferences for distributive justice in the allocation of health care resources: a latent class discrete choice experiment. Medical Decision Making. 2015;35(1):94-105.
83. Spoth R, Ball AD, Klose A, Redmond C. Illustration of a market segmentation technique using family-focused prevention program preference data. Health Education Research. 1996;11(2):259-67.
84. Stanek EJ, Oates MB, McGhan WF, Denofrio D, Loh E. Preferences for treatment outcomes in patients with heart failure: symptoms versus survival. Journal of cardiac failure. 2000;6(3):225-32.
85. Tang C, Xu J, Zhang M. The choice and preference for public-private health care among urban residents in China: evidence from a discrete choice experiment. BMC health services research. 2016;16(1):580.
86. Ungar WJ, Hadioonzadeh A, Najafzadeh M, Tsao NW, Dell S, Lynd LD. Quantifying preferences for asthma control in parents and adolescents using best–worst scaling. Respiratory medicine. 2014;108(6):842-51.
87. van de Wetering L, van Exel J, Bobinac A, Brouwer WB. Valuing QALYs in relation to equity considerations using a discrete choice experiment. PharmacoEconomics. 2015;33(12):1289-300.
88. Veldwijk J, Lambooij MS, Kallenberg FG, van Kranen HJ, Bredenoord AL, Dekker E et al. Preferences for genetic testing for colorectal cancer within a population-based screening program: a discrete choice experiment. European Journal of Human Genetics. 2016;24(3):361-6.
89. Veldwijk J, van der Heide I, Rademakers J, Schuit AJ, de Wit GA, Uiters E et al. Preferences for vaccination: does health literacy make a difference? Medical Decision Making. 2015;35(8):948-58.
90. Virudachalam S, Chung PJ, Faerber JA, Pian TM, Thomas K, Feudtner C. Quantifying parental preferences for interventions designed to improve home food preparation and home food environments during early childhood. Appetite. 2016;98:115-24.
91. Waschbusch DA, Cunningham CE, Pelham Jr WE, Rimas HL, Greiner AR, Gnagy EM et al. A discrete choice conjoint experiment to evaluate parent preferences for treatment of young, medication naive children with ADHD. Journal of Clinical Child & Adolescent Psychology. 2011;40(4):546-61.
92. Weernink MG, van Til JA, van Vugt JP, Movig KL, Groothuis-Oudshoorn CG,

IJzerman MJ. Involving Patients in Weighting Benefits and Harms of Treatment in Parkinson's Disease. PloS one. 2016;11(8):e0160771.
93. Whitty JA, Ratcliffe J, Kendall E, Burton P, Wilson A, Littlejohns P et al. Prioritising patients for bariatric surgery: building public preferences from a discrete choice experiment into public policy. BMJ open. 2015;5(10):e008919.
94. Whitty JA, Stewart S, Carrington MJ, Calderone A, Marwick T, Horowitz JD et al. Patient preferences and willingness-to-pay for a home or clinic based program of chronic heart failure management: findings from the Which? trial. PloS one. 2013;8(3):e58347.
95. Wong Y-N, Egleston BL, Sachdeva K, Eghan N, Pirollo M, Stump TK et al. Cancer patients' trade-offs among efficacy, toxicity and out-of-pocket cost in the curative and non-curative setting. Medical care. 2013;51(9).
96. Yan K, Bridges JF, Augustin S, Laine L, Garcia-Tsao G, Fraenkel L. Factors impacting physicians' decisions to prevent variceal hemorrhage. BMC gastroenterology. 2015;15(1):55.
97. Yoo HI, Doiron D. The use of alternative preference elicitation methods in complex discrete choice experiments. Journal of health economics. 2013;32(6):1166-79.
98. Zimmermann TM, Clouth J, Elosge M, Heurich M, Schneider E, Wilhelm S et al. Patient preferences for outcomes of depression treatment in Germany: a choice-based conjoint analysis study. Journal of affective disorders. 2013;148(2):210-9.

Chapter 3 Explore Preference Heterogeneity For The Treatment of People With Type 2 Diabetes: A Comparison of Random-Parameters And Latent-Class Estimation Techniques

Mo Zhou, MHS, MPA; John F. P. Bridges, PhD

Abstract

There has been an increasing interest in studying patient preference heterogeneity to support regulatory decision-making. While the traditional mixed logit (MXL) and latent class logit (LCL) models have been commonly used to analyze preference heterogeneity in discrete choice data, they have limitations. This study empirically compares an innovative random effect latent class logit (RELCL) model to the traditional approaches using preference data from a discrete-choice experiment among patients with type 2 diabetes. The survey contained 18 pairs of hypothetical diabetes medications that differed in six attributes.

Significant preference heterogeneity was found in all models. The best-fitted RELCL has the lowest BIC (8350.64) and predication error (11.61%) compared to MXL (BIC=8587.38; pred. err.=13.02%) and the best-fitted scale-adjusted LCL (BIC=8403.18; pred. err.=15.69%), indicating improved model fit. Allowing random effect also reduces the number of classes from five in scale-adjusted LCL to two and both have significant policy and clinical implications. RELCL provides the flexibility of LCL and the parsimony of MXL. When significant within-class heterogeneity exists as in patients with prevalent chronic diseases, RELCL may be used to generate more accurate predictions and more parsimonious results that are policy-relevant.

3.1 Introduction

The United States Food and Drug Administration (FDA) issued guidance in 2015 to incorporate patient preference into their regulatory decision-making (CDRH, 2015). Following the issue of guidance, results from two stated-preference studies have been used to support the market approval of new treatments (Ho et al., 2015; Peay et al., 2014). Since then, there has been an increasing interest in studying patient preference because of significant policy implication.

One key regulatory change according to the FDA guidance is that the agency is willing to approve treatments even if the benefit-risk profile is only acceptable to a subset of risk-tolerant patients (CDRH, 2015). Studying preference heterogeneity among patients is therefore meaningful. The demand for studying heterogeneity requires an examination and comparison of different approaches to analyzing preference heterogeneity and guidance on choosing appropriate methods to ensure unbiased results to support regulatory decision-making.

Mixed logit (MXL, also termed random parameter logit) and latent class logit (LCL, also termed finite mixture logit) are the most commonly used models to study unobserved preference heterogeneity in stated-preference literature (McFadden, 1986; Swait, 1994). MXL allows some or all of the preference parameters to be random and vary across individuals according to a continuous probability distribution (Bhat, 1998; Revelt and Train, 1998; Brownstone and Train, 1998). LCL assumes that preferences are distributed discretely into clusters (segments). Preferences are similar among individuals within clusters but vary between clusters (McFadden, 1986). Heterogeneity in variances (i.e. scales) is sometimes adjusted in addition to preference heterogeneity in a scale-adjusted latent class logit (SLCL) model (Fiebig et al, 2010; Louviere et al, 2000). SLCL allows respondents in the same preference group to have different variances (i.e. scale parameters) that reflect the consistency of their choices compared to preference (Flynn et al, 2010; Magidson and Vermunt, 2007; Campbell et al, 2011).

Many stated-preference studies in health use LCL or SLCL to analyze preference heterogeneity because the segmentation results have more straightforward interpretation than MXL. Despite weaker assumptions on the distributions of parameters, LCL often fits the data at least as well as MXL (Louviere, 2006). However, one limitation of LCL or SLCL model is that it assumes preference coefficients take only a set of distinct values and are homogeneous within classes, which sometimes does not hold empirically. When there is substantial within-class preference heterogeneity or significant overlap between classes, LCL often leads to too many classes (Lenk and DeSarbo, 2000). As MXL can accommodate more extensive heterogeneity with fewer parameters, it may not be sufficient when there are sizeable subgroups.

As a remedy, Lenk and DeSarbo (2000) proposed using random effects in LCL models, which provides the flexibility of LCL and the parsimony of MXL. Given the complication and wide variation of preference in healthcare, random effect LCL (RELCL) may better capture the preference distribution. However, few stated-preference studies have tested this approach. In this paper, we sought to empirically compare RELCL to the traditional MXL and LCL/SLCL in analyzing unobserved preference heterogeneity. We also demonstrate the approaches to testing the underlying distribution of preferences and selecting the appropriate model to analyze preference heterogeneity for researchers who are interested in studying patient preference.

The data is from a discrete-choice experiment (DCE) survey among patients with type 2 diabetes on preference for diabetes medications. Preference for treatment for type 2 diabetes is selected as a case study because diabetes is a prevalent chronic condition that affects nearly 30 million individuals in the US and the complications of diabetes result in significant morbidity and costs (CDC, 2014). While treatment options have expanded (Tran et al, 2015), poor adherence to known effective interventions persists. Understanding patient preference and incorporating it in drug development and clinical guidelines could therefore improve adherence, satisfaction, and quality of life, and have significant public health impact.

3.2 Methods

3.2.1 DCE survey instrument

A robust, comprehensive, and engaged process was used to develop the survey (Janssen et al, 2016). The attributes in the survey included the amount of reduction in hemoglobin A1c, the duration of stable blood glucose levels, frequency of hypoglycemia, duration of nausea per day, treatment burden, and cost (Table 3-1). Each attribute had three levels. Information about the attributes and levels was provided to respondents before they answered the DCE choice tasks. The attributes and levels as well as the framing of the information was pretested among patients to make sure they were appropriate.

[TABLE 3-1 INSERT HERE]

A Bayesian D-efficient design generated by Ngene was used, with priors obtained from the pilot data. The design consisted of 48 choice tasks divided into three blocks. The blocks were selected to minimize the correlation between blocks and attribute levels.

Each block contained 16 choice tasks plus two additional holdout tasks. One holdout task repeated one of the 16 tasks in the block to assess test-retest reliability. The other one remained the same across all three blocks to test systematic difference in preference between blocks. Each respondent was randomly assigned to a block. Choice tasks were presented in random order. Each task asked respondents to choose between two hypothetical mediations for type 2 diabetes (Figure 3-1). The survey also asked the respondents' diabetes history, current disease management, health status, health behaviors (e.g. smoking, exercise), and personalities measured by Likert scales.

[FIGURE 3-1 INSERT HERE]

3.2.2 Data collection

A national survey was conducted among patients with type 2 diabetes through the GfK KnowledgePanel in 2015. It is an online panel that provides sampling coverage of 97% of the US adult population (Couper, 2000). GfK surveys include both listed and unlisted numbers, and are not limited to current Internet users or computer owners. There has been shown to be over 4,000 potential respondents with type 2 diabetes in the panel, with a high proportion of complex patients (Safford et al, 2007). Due to higher prevalence of diabetes (CDC, 2014) and lower adherence rates to treatments (Egede et al, 2011), African Americans and Hispanics were oversampled in this study. Hispanics had the option to take the survey in English or Spanish. GfK provided respondents' socio demographic data and calculated sample weights to rebalance the sample to national norms for the assessment of aggregate results. This study has been exempt from the Johns Hopkins School of Public Health IRB.

3.2.3 Statistical analysis

The DCE data was analyzed in a random utility framework (McFadden, 1974), where individuals are assumed to choose the option with the greatest utility. The utility that person n, n = 1,…,N, obtains from choosing medicine i, i = 1,2, in choice task t, t = 1,…,16, can be written as:

$$U_{nit} = x'_{it}\beta + \varepsilon_{nit} \quad [1]$$

where x is a vector of the six attributes of medicine i, β is a vector of utility values for these attributes, and ε is an independently and identically distributed (*iid*) error term. When ε follows a type I extreme value distribution (also termed Gumbel distribution), the probability of person n choosing medicine i over j in choice task t is:

$$P_n\left(i_t \middle| \beta, x_t\right) = \frac{\exp\left(x'_{it}\beta\right)}{\exp\left(x'_{it}\beta\right) + \exp\left(x'_{jt}\beta\right)} = \frac{1}{1 + \exp\left(x_{jt} - x_{it}\right)'\beta}. \quad [2]$$

The probability of individual n choosing a sequence of choices in the 16 choice tasks is then:

$$P_n\left(y \middle| \beta, x\right) = \prod_{t=1}^{16} P_n\left(i_t \middle| \beta, x_t\right) \quad [3]$$

where y represents a vector of choices that respondent n makes in the 16 choice tasks.

While interaction term can be added between β and individual characteristics (e.g. age, gender) to capture preference heterogeneity between groups, the model in equation [2] (i.e., conditional logit, CL) assumes no unobserved preference heterogeneity among respondents. The assumption that preference heterogeneity can be accurately determined *a priori* by observed variables often does not hold empirically (Morey and Greer Rossmann, 2003; Iraguen and de Dios Ortuzar, 2004).

To analyze unobserved preference heterogeneity, we first used MXL, assuming the preference weights (β) for all attribute levels are normally distributed across respondents. The preference weights of individual n are then specified as:

$$\beta_n = \beta + \sigma u_n \qquad [4]$$

where β is a vector of mean preference weights for the attribute levels, u_n is multivariate normally distributed with correlation matrix R, and σ is a vector of standard deviations of the preference weights. We assumed independent preference weights for different levels of the same attribute and allowed them to have different variances. When $\sigma = 0$, MXL reduces to CL.

MXL reports weighted average β across all respondents. Based on these estimates, we calculated the relative attribute importance (RAI) for each attribute by dividing the difference between the maximum and minimum preference weights for the levels within an attribute by the sum of such differences of all six attributes. RAI measures the relative weight that respondents give to an attribute when they make choices.

We then assumed preference follows a discrete distribution, with each cluster q, $q = 1,\ldots,Q$, having a class-specific preference β^q that varies between clusters (McFadden, 1986). The probability of a sequence of choices from individual n is then:

$$P_n\left(y \middle| \beta, x\right) = \sum_{q=1}^{Q} \pi_{q|n} P_n\left(y \middle| \beta^q, x\right). \qquad [5]$$

where $\pi_{q|n}$ is the probability of individual n belonging to class q that is estimated simultaneously with the choice model based on the choice patterns (Hole, 2008).

In order to adjust for potential heterogeneity in variances, we allowed respondents in the same preference group to have different scale parameters (Flynn et al, 2010;

Magidson and Vermunt, 2007; Campbell et al, 2011). The probability of person n in class q choosing medicine i over j in choice task t can be expressed as:

$$P_n\left(i_t \middle| \beta^q, x_t\right) = \sum_{s=1}^{S} \pi_{s|n} \left(\frac{1}{1+\exp\left(\lambda_s \left(x_{jt} - x_{it}\right)' \beta^q\right)} \right) \qquad [6]$$

where $\pi_{s|n}$ is the probability of individual n belonging to scale class s, $s = 1,\ldots,S$, and λ_s, $\lambda_s \in (0,1)$, is the scale parameter for scale class s that is inversely correlated with the variance of choices. The higher the scale gets, the smaller the variance is and the higher the certainty level is. To achieve identification, λ_1 (i.e., the parameter for scale class 1) is normalized to 1.

We estimated SLCL with two to six preference classes and one to three scale classes. Given that Akaike Information Criterion (AIC) may overestimate the number of classes (Celeux and Soromenho 1996) and that Bayesian Information Criterion (BIC) has been shown to have better performance (Nylund et al, 2007), minimized BIC was used to identify the appropriate number of classes. Preference heterogeneity for each attribute between classes was tested using Wald test. RAI was calculated and compared across classes.

Finally, we allowed within-class heterogeneity in LCL model by adding random effects. We assumed that preference weights for all attribute levels were normally distributed with different variances within each class. The preference weights for individual n in class q are then:

$$\beta_n^q = \beta^q + \sigma^q u_n^q \qquad [7]$$

where σ^q is a vector of standard deviations of preference weights for all attribute levels in class q, and u_n^q is multivariate normally distributed. The model reduces to LCL when $\sigma =$

0 for all q, and is the traditional MXL when $Q = 1$. We estimated RELCL with two to six classes, and selected the best-fitted model based on BIC. RAI was calculated and compared across classes.

We compared MXL, LCL, SLCL, and RELCL based on BIC, prediction error, and class assignment. To examine the individual characteristics that are associated with class membership, demographics, socioeconomic, and health status variables were incorporated in the best-fitted model by specifying the class membership probability, $\pi_{q|n}$, via a multinomial link function with these characteristics. Backward selection was used to determine which variables to include in the final model.

Analysis was conducted in Latent Gold Choice 5.1, which combines Expectation-Maximization and Newton-Raphson algorithms to estimate maximum likelihood solutions of finite mixture models. Because the estimation may sometimes reach local maxima that are not representative (Lanza and Rhoades 2011; Nylund et al. 2007), each model was replicated 10 times with random starting seeds to look for global maxima. Because the probability of choosing a medicine depends only on the differences in levels of medicine attributes between two medicines, the first level of each attribute is usually normalized to zero. In this study, we used effect-coding to achieve identification. Sampling weights were taken into account when calculating standard errors, the latent class distribution, and covariate effects.

3.3 Results

552 respondents with type 2 diabetes filled out the survey. 543 completed all choice tasks and were included in the analysis. The mean age was 61 years. 49% were female. The sample has higher percentage of African Americans (23%) and Hispanics

(22%) than the general population due to oversampling. Respondents in the sample were also more educated and had higher income than the general population. The average type 2 diabetes history was 11 years. 90% of respondents measured hemoglobin A1c at least once in the past six months. 41% had A1c below 7 percent as recommended by clinical guideline. Only 7% were not taking any diabetes treatments. 59% of the sample had no complications. Majority (77%) reported good, very good, or excellent overall health. Most respondents were optimistic about the future (68%) and had been actively improving health (69%). (Table 3-2)

[TABLE 3-2 INSERT HERE]

All attributes had significant impacts on choices in MXL model ($p<0.001$) (Table 3-3). Most effects were in the expected direction. Larger reduction in hemoglobin A1c, longer duration of stable blood glucose level, fewer hypoglycemia events, less nausea, and lower cost were associated with higher preference weights. In terms of treatment burden, respondents were indifferent between having one pill a day (0.334) and having two pills a day (0.374) ($p>0.05$), but had a significantly lower preference weight for having injection (-0.708). The standard deviations for all attributes but frequency of hypoglycemia ($p=0.089$) were statistically significant ($p<0.05$), indicating significant preference heterogeneity for the five attributes among respondents. Duration of nausea was considered the most important attribute (RAI=0.272), while frequency of hypoglycemia was the least important overall for choices (RAI=0.086).

[TABLE 3-3 INSERT HERE]

The best-fitted SLCL model contained five preference classes and two scale classes. Allowing two scale classes improved BIC from 8440.30 to 8403.18, indicating heterogeneity in scale among respondents. There were significant preference heterogeneity for reduction in hemoglobin A1c ($p<0.001$), duration of stable blood glucose level ($p=0.001$), and treatment burden ($p=0.011$) across classes (Table 3-4).

[TABLE 3-4 INSERT HERE]

Each class in the 5-class SLCL model focuses predominantly on one attribute while making choices (Figure 3-2). Class 1 (cost dominant class, 24.0% of sample) considered cost the most important attribute (RAI=0.47); class 2 (treatment burden dominant class, 23.7%) focused on treatment burden (RAI=0.42) while gave little consideration to effectiveness (RAI=0.04); class 3 (nausea dominant class, 22.5%) concerned mainly about duration of nausea (RAI=0.56); class 4 (effectiveness dominant class, 18.4%) cared most about the amount of reduction in hemoglobin A1c (RAI=0.48); and class 5 (effect stability dominant class, 11.4%) made choices largely based on duration of stable blood glucose level (RAI=0.43). Overall, 42% of respondents were in scale class 2 and were less consistent with their choices ($\lambda=0.339$).

[FIGURE 3-2 INSERT HERE]

Two classes were identified when random effects were allowed in LCL. Preferences for medicine effectiveness ($p=0.002$), duration of nausea ($p<0.001$), and treatment burden ($p<0.001$) were significantly different between the classes (Table 3-5). Significant within-class preference heterogeneity was found for all attributes in both classes ($p<0.05$), except the preference for frequency of hypoglycemia in class 2 ($p>0.05$).

[TABLE 3-5 INSERT HERE]

Figure 3-3 shows that class 1 (cost-effectiveness class, 55.9% of sample) in the RELCL model considered cost (RAI=0.26) and the amount of reduction in hemoglobin A1c (RAI=0.25) the most important attributes when choosing diabetes medicines, and treatment burden the least important attribute (RAI=0.08). Class 2 (side-effect focused class, 44.1%) mainly focused on the duration of nausea (RAI=0.37) and treatment burden (RAI=0.28) when making choices and valued the effectiveness of medicine less than class 1.

[FIGURE 3-3 INSERT HERE]

Given that MXL has higher BIC (8587.38) than LCL (8438.82), SLCL (8403.18), and RELCL (8350.64), there is likely to be clusters of preferences in the data (Table 3-6). Comparing among the discrete models, allowing random effect in LCL dramatically reduces the number of classes and increases degree of freedom, despite the additional preference variance estimates in each class. In terms of model performance, allowing within-class heterogeneity reduces BIC and prediction error (from 15.76% in LCL and 15.69% in SLCL to 11.61% in RELCL), indicating improved model fit.

[TABLE 3-6 INSERT HERE]

Including relevant individual characteristics in the class membership probability function in the 2-class RELCL model further reduces BIC (8319.85) and model predictor error (11.37%) and improves model fit. Comparing the segmentation results from this model to those from SLCL, RELCL combines the cost dominant class (i.e., class 1), effectiveness dominant class (i.e., class 4) and most of effect stability dominant class (i.e., class 5) in SLCL into one class (i.e., cost-effectiveness class), and the treatment burden dominant class (i.e., class 2) and nausea dominant class (i.e., class 3) in SLCL into another class (i.e., side effect focused class) due to proximity of preferences between classes (Table 3-7).

[TABLE 3-7 INSERT HERE]

Respondents' age, income level, self-reported health status, current diabetes treatment, and perspective about the future were correlated with class membership in the 2-class RELCL model. A 10-year increase in age was associated with a 0.25 reduction in the odds of being in class 1 ($p<0.001$). Individuals with under $25,000 annual income had 0.42 higher odds of being in class 1 than the general average ($p=0.004$). Individuals who reported excellent health status had 0.78 higher odds of being in class 1 than general average ($p=0.005$). Currently using insulin injections or insulin injections plus oral medicines for diabetes treatment was associated with a 0.74 ($p=0.004$) and 0.63 ($p<0.001$) increase, respectively, in the odds of being in class 1, while having no current diabetes treatments was associated with a 1.39 increase in the odds of being in class 2 ($p<0.001$). Respondents who reported being optimal about future had 0.27 higher odds to be in class 1 ($p=0.022$). (Table 3-5)

3.4 Discussion

This study demonstrates the application of RELCL in modeling unobserved preference heterogeneity in the context of health. By comparing various model specifications, we tested the assumptions on the underlying distribution of preferences. In our DCE data, we first rejected the hypothesis that preference is normally distributed across the respondents and found distinct preference types. By allowing random effect, we then found two main preference classes with some degree of preference variation within each class that is better modeled by a normal distribution than additional classes.

Given that type 2 diabetes is a prevalent chronic condition that occurs in a wide range of population, patient preference is diverse and difficult to be categorized into distinct groups. With more flexible model specification, RELCL better captures this complex preference pattern than LCL, where within-class homogeneous preference is assumed, leading to improved model fit and prediction accuracy. When the patient group is less diverse as in rare diseases or patient preference is less heterogeneous as in acute and severe conditions, random effect may not be necessary and LCL may be more appropriate to describe the preference patterns. That being said, RELCL can still serve as a tool to test the underlying preference distribution in preference studies in health to ensure unbiased estimates to support regulatory decision-making.

RELCL can reduce over-fitting of data. When there is significant within-class heterogeneity or overlap between classes, LCL may generate more classes than desired to fit the data. Despite better fit, the result is difficult to use by decision-makers. This study shows that RELCL combines overlapped preference types and describes some of the preference variations using deviation from the mean on a continuous distribution instead of additional class. RELCL therefore generates more parsimonious segments than LCL. Given that regulatory and clinical decision-makers often prefer parsimonious results due to limited financial resources and capacity to accommodate many preference types, RELCL can be used to produce more practical results for policy and clinical decision-making.

This study found significant preference heterogeneity among patients with type 2 diabetes. Patients with younger age, lower income, excellent overall health status, optimistic view about future or currently using injection to treat diabetes focused on cost and effectiveness when choosing medicine. Patients who were older, have higher income, worst overall health status, or not taking any diabetes treatment were more likely to make medicine choices based on side effects and treatment burden. These results are consistent with some studies in the current literature. For example, Casciano et al (2011) also found that administration mode (oral versus injection) was the most important factor when diabetic patients chose treatments, but those who were already treated with insulin considered side effects, maintenance of blood sugar levels, and risk of hypoglycemia to be more important than administration mode. Our findings on the correlation between age, income, health status, personality and treatment preference, especially after controlling for potential confounding individual factors, complement the current literature on patient preference heterogeneity over diabetes treatments.

The results have significant policy and clinical implications. Cost-effective medicines should be provided and recommended to the younger, lower income, and relatively healthier group and people who are currently taking injections. Additional pills or injection can be used to reduce cost and improve effectiveness if necessary. The older, wealthier, and sicker patients as well as those who just start treating diabetes should have access to medicines with fewer side effects, even for a higher price or less effectiveness. FDA could take such preference into consideration during the drug approval process to accommodate different patient needs, and clinicians can choose treatment based on the preference types. Given that diabetes requires meaningful patient engagement to improve outcomes, tailored treatment will improve adherence and health outcome.

This study has several limitations. First, it is likely to be underpowered to detect more classes. Increasing the number of classes in RELCL generated more separated classes (measured by Entropy R^2) and reduced the prediction error. However, because the increased number of parameters associated with additional classes dramatically reduced degree of freedom, BIC decreased. A larger sample size may allow more parameters and more classes. Scale was not adjusted in RELCL to reduce the number of parameters, but it is partially incorporated in the within-class preference heterogeneity.

Second, the true preference heterogeneity distribution is unknown in this empirical data, and therefore we were unable to assess how well each model reveals the true distribution parameters except using BIC. Future study should use simulation to theoretically test the performance of RELCL against LCL and explore under what circumstances (e.g. the level of overlap between classes) RELCL performs better than LCL. Finally, the sample in this study had higher income and more education than the general diabetes patient population. Although sampling weights were used to adjust standard errors and latent class distribution, the lower socioeconomic group could have different preference over medicine that was underrepresented in the sample. Extra caution should be used to generalize the results to the entire diabetes population.

Despite limitations, this study demonstrated significant preference heterogeneity among patients with type 2 diabetes and linked the preference types to some individual characteristics. The results can be used to inform patient-centered drug development as well as personalized clinical care to improve patient satisfaction and treatment adherence. This study also tested several model specifications and demonstrated the approaches to studying unobserved preference heterogeneity. It can serve as guidance for researchers who are interested in studying preference heterogeneity in health.

Reference

Bhat CR. Accommodating flexible substitution patterns in multi-dimensional choice modeling: Formulation and application to travel mode and departure time choice. Transportation Research Part B: Methodological. 1998;32(7):455-66.

Brownstone D and Train K. Forecasting new product penetration with flexible substitution patterns. Journal of Econometrics. 1998;89(1-2):109-29.

Campbell D, Hensher DA, and Scarpa R. Non-attendance to attributes in environmental choice analysis: a latent class specification. J Environ Plann Manag. 2011;54(8):1061–76.

Casciano R, Malangone E, Ramachandran J, et al. A quantitative assessment of patient barriers to insulin. International Journal of Clinical Practice. 2011;65(4):408–14.

Celeux G and Soromenho G. An entropy criterion for assessing the number of clusters in a mixture model. Journal of Classification. 1996;13:195–212.

Center for Devices and Radiological Health (CDRH) and Center for Biologics Evaluation and Research (CBER). Patient preference information – submission, review in PMAs, HDE applications, and *de novo* requests, and inclusion in device labeling: draft guidance for industry, Food and Drug Administration staff, and other stakeholders. Issued on May 18, 2015.

Centers for Disease Control and Prevention. National Diabetes Statistics Report: Estimates of Diabetes and Its Burden in the United States, 2014. Atlanta, GA: U.S. Department of Health and Human Services; 2014.

Couper MP. Review: web surveys: a review of issues and approaches. The Public Opinion Quarterly. 2000;64(4):464-94.

Egede LE, Gebregziabher M, Hunt KJ, et al. Regional, geographic, and ethnic differences in medication adherence among adults with type 2 diabetes. The Annals of Pharmacotherapy. 2011;45(2):169-78.

Fiebig D, Keane M, Louviere J, et al. The generalized multinomial logit model: Accounting for scale and coefficient heterogeneity. Mark Sci. 2010;29(3):393–421.

Flynn TN, Louviere JJ, Peters TJ, et al. Using discrete choice experiments to understand preferences for quality of life. Variance-scale heterogeneity matters. Social Science and Medicine. 2010;70(12):1957–65.

Ho MP, Gonzalez JM, Lerner HP, et al. Incorporating patient-preference evidence into regulatory decision making. Surgical Endoscopy. 2015;29(10):2984-93.

Hole AR. Modelling heterogeneity in patients' preferences for the attributes of a general practitioner appointment. Journal of Health Economics. 2008;27(4):1078-94.

Iraguen P, and de Dios Ortuzar J. Willingness-to-pay for reducing fatal accident risk in urban areas: an internet-based web page stated preference survey. Accident Analysis & Prevention. 2004;36(4):513-24.

Janssen EM, Segal JB, and Bridges JF. A framework for instrument development of a choice experiment: an application to type 2 diabetes. Patient. 2016;9(5):465-79.

Lanza ST, and Rhoades BL. Latent class analysis: An alternative perspective on subgroup analysis in prevention and treatment. Prevention Science. 2011;1–12.

Lenk PJ, and DeSarbo WS. Bayesian inference for finite mixture models of generalized linear models with random effect. Psychometrika. 2000;65: 93-119.

Louviere JJ, Hensher DA, and Swait JD. Stated Choice Methods: Analysis and Applications. Cambridge, MA: Cambridge University Press; 2000.

Louviere JL. What you don't know might hurt you: some unresolved issues in the design and analysis of discrete choice experiments. Environmental Resource Economics. 2006;34:173–88.

Magidson J, and Vermunt JK. Removing the scale factor confound in multinomial logit choice models to obtain better estimates of preference. Sawtooth Software Conference. 2007;1–18.

McFadden D. Conditional logit analysis of qualitative choice behavior. In: P. Zarembka, editor. Frontiers in econometrics. New York: Academic Press; 1974.

McFadden D. The choice theory approach to market research. Marketing Science. 1986;5(4):275-97.

Morey E and Greer Rossmann K. Using stated-preference questions to investigate variations in willingness to pay for preserving marble monuments: Classic heterogeneity, random parameters, and mixture models. Journal of Cultural Economics. 2003;27(3):215-29.

Nylund KL, Asparouhov T, and Muthén BO. Deciding on the number of classes in latent class analysis and growth mixture modeling: a Monte Carlo simulation study. Structural Equation Modeling. 2007;14:535–569.

Peay HL, Hollin I, Fischer R, et al. A community-engaged approach to quantifying caregiver preferences for the benefits and risks of emerging therapies for Duchenne muscular dystrophy. Clinical Therapeutics. 2014;36(5):624-37.

Revelt D and Train K. Mixed logit with repeated choices: households' choices of appliance efficiency level. Review of Economics and Statistics. 1998;80(4):647-57.

Safford MM, Allison JJ, Kiefe CI. Patient complexity: More than comorbidity. The vector model of complexity. Journal of General Internal Medicine. 2007;22:382-90.

Swait J. A structural equation model of latent segmentation and product choice for cross-sectional revealed preference choice data. Journal of Retailing and Consumer Services. 1994;1(2):77-89.

Tran L, Zielinski A, Roach AH, et al. Pharmacologic treatment of type 2 diabetes: oral medications. The Annuals of Pharmacotherapy. 2015;49(5):540-56.

Figure 3-1 Sample choice task in the discrete choice experiment survey

Consider the following two diabetes medicines. Which medicine would you prefer?

Select <u>one choice</u> – either Medicine A or Medicine B at the bottom of the list.

Attributes	Medicine A	Medicine B
A1c levels go down by	1%	0.5%
Stable blood glucose	2 days per week	4 days per week
Low blood glucose	During the day	None
Nausea	None	30 minutes per day
Additional medicine	2 pills per day	1 pill per day
Additional out-of-pocket costs	$50 per month	$30 per month
Which medication would you choose? (pick one)	**I would choose Medicine A** ☐	**I would choose Medicine B** ☐

Figure 3-2 Relative attribute importance from the scale-adjusted latent class model

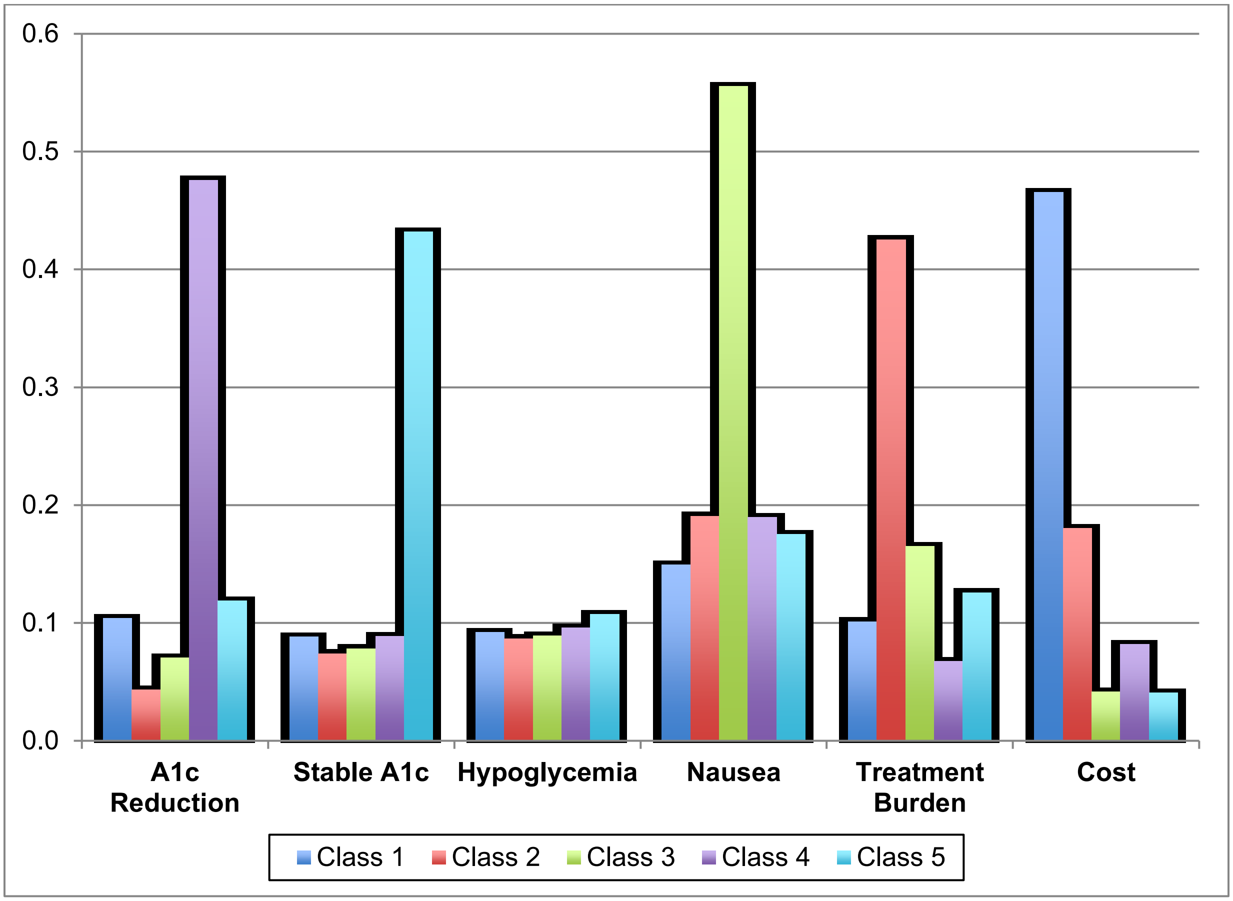


Figure 3-3 Relative attribute importance from the random effect latent class logit model

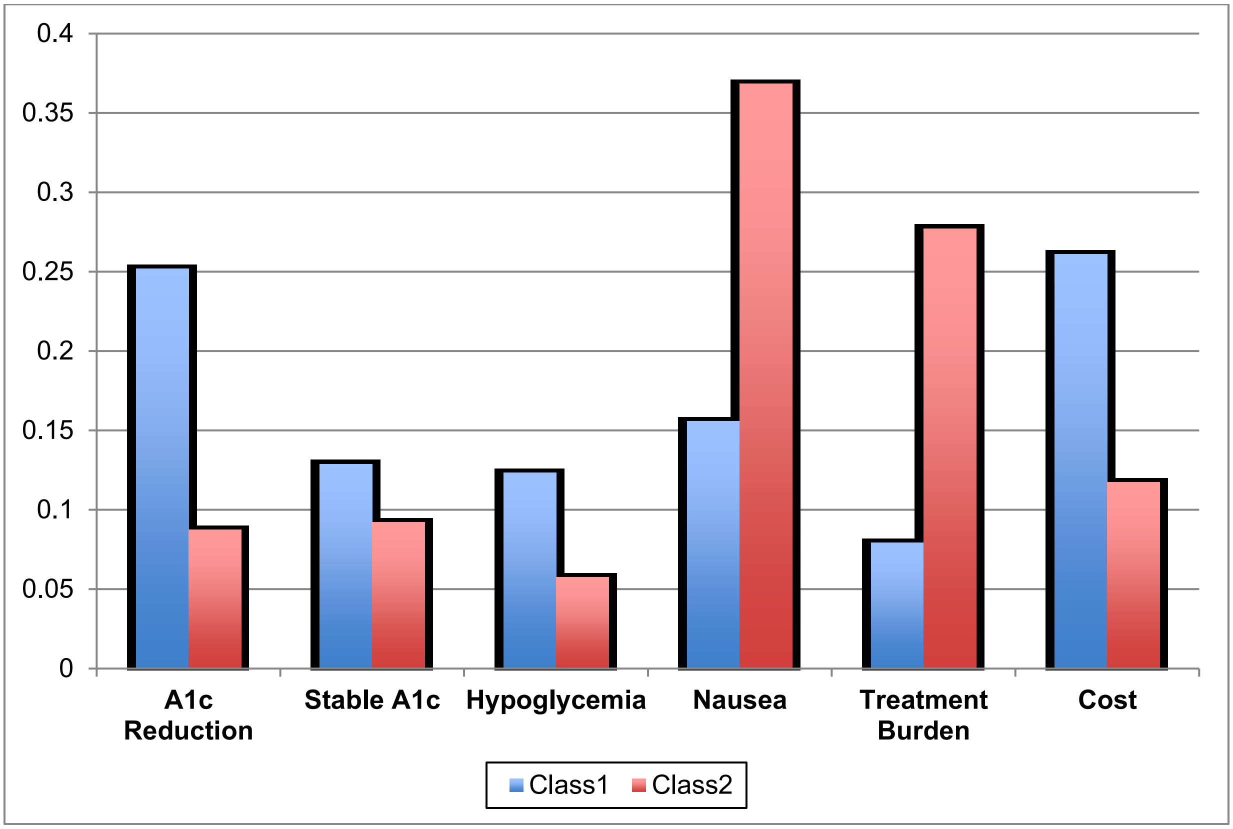


Table 3-1 Attributes and levels included in the discrete choice experiment survey

Attribute	Levels
Reduction in hemoglobin A1c level	• 1% • 0.5% • 0%
Duration of blood glucose level remaining stable	• 6 days per week • 4 days per week • 2 days per week
Frequency of hypoglycemia	• None • During the day only • During the day and/or at night
Duration of nausea	• None • 30 minutes per day • 90 minutes per day
Treatment burden in addition to current medicine	• 1 pill per day • 2 pills per day • 1 pill and 1 injection per day
Out-of-pocket cost	• $10 per month • $30 per month • $50 per month

Table 3-2 Respondent characteristics

	N	Percent
Demographic Characteristics		
Age (mean)	552	61.30
Gender		
Male	279	0.51
Female	273	0.49
Race		
White	289	0.52
Black	126	0.23
Hispanic	119	0.22
Other	18	0.03
Education		
Less than high school	43	0.08
High school	188	0.34
Some college	156	0.28
Bachelor's degree or higher	165	0.30
Income		
< 25,000	132	.24
25,000 – 50,000	148	.27
50,000 – 74,999	111	.20
≥ 75,000	161	.29
Health-related Characteristics		
Years of diagnosis (mean)	545	11.24
Self-Reported Health		
Excellent	33	.06
Very good	158	.29
Good	232	.42
Fair	106	.19
Poor	23	.04
No. A1c measured in last 6 months		
None	41	.07
1 time	239	.43
> 2 times	261	.47
Don't know	10	.02
Most recent A1c level		
≥ 7.0%	233	.43
< 7.0%	228	.41
Don't know	89	.16
Type of diabetes medicine used		
No prescription medicine	37	.07
Only pills	345	.63
Only injections/shots	42	.08
Pills and injections/shots	127	.23
No complications	326	.59
No other chronic conditions	92	.17
Personality Characteristics		
Agree or strongly agree with the following statement		
I'm always optimistic about my future	372	.68
I have a lot of self-control	316	.57
I'm actively working to improve my health	375	.69
I consider myself a risk-taker	98	.18

Table 3-3 Estimates from the mixed logit model

Attributes	Coefficient (SE)	p-value[a]	Std. Dev. (SE)	p-value[a]	RAI[b]
A1c reduction					
1%	0.364*** (.056)	<.001	0.488*** (.078)	<.001	.162
0.50%	0.218*** (.049)		0.103** (.040)		
0%	-0.581*** (.057)		0.591*** (.083)		
Stable glucose					
6 days/week	0.283*** (.044)	<.001	0.176** (.055)	.003	.107
4 days/week	0.060 (.047)		0.078* (.039)		
2 days/week	-0.343*** (.042)		0.098 (.057)		
Hypoglycemia					
None	0.193*** (.038)	<.001	0.099* (.046)	.089	.086
Day	0.115* (.046)		0.021 (.040)		
Day & night	-0.308*** (.038)		0.078 (.049)		
Nausea					
None	0.704*** (.063)	<.001	0.787*** (.067)	<.001	.272
30 min/day	0.180*** (.038)		0.090* (.044)		
90 min/day	-0.885*** (.064)		0.877*** (.069)		
Treatment burden					
1 pill	0.334*** (.049)	<.001	0.375*** (.056)	<.001	.185
2 pills	0.374*** (.055)		0.256*** (.057)		
1 pill & 1 injection	-0.708*** (.059)		0.631*** (.092)		
Cost					
$10/month	0.526*** (.084)	<.001	0.839*** (.093)	<.001	.189
$30/month	0.054 (.097)		0.093 (.100)		
$50/month	-0.580*** (.089)		0.746*** (.074)		

a. P-values were from the Wald-test. b. RAI: relative attribute importance.
* $p<0.05$ ** $p<0.01$ *** $p<0.001$

Table 3-4 Estimates from the 5-class 2-scale-class latent class logit model

Attributes	Coefficient (SE)					Wald(=)[a]
	Class 1	Class 2	Class 3	Class 4	Class 5	p-value
A1c reduction						
1%	0.338	0.226	0.189	1.564***	0.218	<.001
	(.257)	(.120)	(.123)	(.347)	(.154)	
0.50%	0.343	-0.135	0.170	0.531*	0.211	
	(.189)	(.208)	(.288)	(.212)	(.170)	
0%	-0.681***	-0.091	-0.359	-2.096***	-0.429**	
	(.166)	(.155)	(.324)	(.450)	(.136)	
Stable glucose						
6 days/week	0.370	0.386	0.250*	0.447	1.010***	.001
	(.275)	(.201)	(.114)	(.242)	(.235)	
4 days/week	0.128	-0.156	0.109	-0.236	0.330*	
	(.238)	(.114)	(.129)	(.247)	(.140)	
2 days/week	-0.498**	-0.230	-0.359*	-0.211	-1.339***	
	(.166)	(.155)	(.147)	(.123)	(.264)	
Hypoglycemia						
None	0.322	0.399**	0.106	0.184	0.336*	.550
	(.189)	(.132)	(.286)	(.136)	(.154)	
Day	0.262	-0.074	0.293	0.278	-0.089	
	(.159)	(.114)	(.510)	(.191)	(.201)	
Day & night	-0.584***	-0.325**	-0.399	-0.461**	-0.247	
	(.142)	(.102)	(.277)	(.147)	(.188)	
Nausea						
None	0.598*	0.845*	1.930*	0.672**	0.157	.270
	(.285)	(.337)	(.865)	(.242)	(.153)	
30 min/day	0.278**	-0.098	0.456	0.116	0.398	
	(.094)	(.152)	(.302)	(.145)	(.273)	
90 min/day	-0.875**	-0.748**	-2.387*	-0.788***	-0.555**	
	(.285)	(.268)	(1.137)	(.219)	(.187)	
Treatment burden						
1 pill	0.519	1.303***	0.256	0.006	0.187	.011
	(.294)	(.379)	(.164)	(.156)	(.141)	
2 pills	-0.040	0.947***	0.513	0.257	0.249	
	(.248)	(.226)	(.549)	(.225)	(.163)	
1 pill & 1 injection	-0.478***	-2.250***	-0.769	-0.263	-0.436**	
	(.140)	(.560)	(.603)	(.219)	(.137)	
Cost						
$10/month	2.442**	0.614*	0.116	0.242	0.104	.130
	(.769)	(.306)	(.243)	(.326)	(.273)	
$30/month	-0.295	0.277	-0.209	0.148	-0.118	
	(.317)	(.232)	(.368)	(.601)	(.533)	
$50/month	-2.147**	-0.891***	0.093	-0.390	0.014	
	(.680)	(.231)	(.417)	(.418)	(.410)	
Class Share	0.240	0.237	0.225	0.184	0.114	

a. Wald test for equal impact of each attribute across classes.
*p<0.05 **p<0.01 ***p<0.001

Table 3-5 Estimates from the random effect latent class logit model with individual covariates as class membership predictors

	Class 1 (cost-effectiveness)		Class 2 (side-effect focused)		Wald(=)[a] p-value	
	Coef. (SE)	Std. Dev.	Coef. (SE)	Std. Dev.	Coef. (SE)	Std. Dev.
<u>**Attributes**</u>						
A1c reduction						
1%	0.583*** (.096)	0.847*** (.090)	0.337*** (.086)	0.192 (.170)	0.002	<.001
0.50%	0.317*** (.070)	0.064 (.065)	0.028 (.101)	0.181* (.086)		
0%	-0.899*** (.105)	0.911*** (.102)	-0.365*** (.092)	0.012 (.156)		
Stable glucose						
6 days/week	0.370*** (.067)	0.019 (.089)	0.361*** (.086)	0.376*** (.093)	1.000	<.001
4 days/week	0.017 (.071)	0.185*** (.053)	0.020 (.080)	0.172 (.130)		
2 days/week	-0.387*** (.064)	0.166* (.081)	-0.381*** (.093)	0.548*** (.111)		
Hypoglycemia						
None	0.283*** (.065)	0.063 (.088)	0.221** (.085)	0.029 (.081)	0.170	0.001
Day	0.160* (.070)	0.206*** (.060)	0.019 (.090)	0.026 (.086)		
Day & night	-0.443*** (.062)	0.270** (.085)	-0.204** (.078)	0.055 (.075)		
Nausea						
None	0.361*** (.072)	0.420*** (.078)	1.355*** (.138)	0.872*** (.115)	<.001	<.001
30 min/day	0.195*** (.055)	0.063 (.063)	0.249** (.083)	0.176* (.082)		
90 min/day	-0.556*** (.077)	0.357*** (.099)	-1.604*** (.151)	1.049*** (.142)		
Treatment burden						
1 pill	0.161* (.067)	0.012 (.074)	0.760*** (.111)	0.592*** (.107)	<.001	<.001
2 pills	0.144 (.078)	0.214** (.082)	0.708*** (.123)	0.470*** (.086)		
1 pill & 1 injection	-0.305*** (.061)	0.203* (.089)	-1.468*** (.154)	1.061*** (.132)		
Cost						
$10/month	0.788*** (.140)	1.047*** (.138)	0.401* (.170)	0.538*** (.151)	0.280	<.001
$30/month	-0.042 (.135)	0.165 (.112)	0.140 (.193)	0.054 (.179)		
$50/month	-0.747*** (.134)	0.882*** (.137)	-0.542*** (.149)	0.591*** (.147)		

Individual Covariates [b]

	Coef. (SE)	Odds Ratio	Coef. (SE)	Odds Ratio	Wald p-value
Age	-0.029*** (.007)	0.972***	0.029*** (.007)	1.029***	<.001
Income					
< $25,000	0.352** (.133)	1.422**	-0.352** (.133)	0.703**	0.027
$25,000-49,999	-0.061 (.132)	0.941	0.061 (.132)	1.063	
$50,000-74,999	-0.307* (.132)	0.736*	0.307* (.132)	1.359*	
≥ $75,000	0.015 (.114)	1.015	-0.015 (.114)	0.985	
Health status					
Excellent	0.577* (.227)	1.781*	-0.577* (.227)	0.561*	0.046
Very good	-0.225 (.146)	0.799	0.225 (.146)	1.252	
Good	-0.058 (.120)	0.944	0.058 (.120)	1.059	
Fair	0.112 (.164)	1.119	-0.112 (.164)	0.894	
Poor	-0.407 (.229)	0.666	0.407 (.229)	1.502	
Current treatment					
No treatment	-0.869*** (.267)	0.419***	0.869*** (.267)	2.385***	<.001
Only pills	-0.175 (.135)	0.839	0.175 (.135)	1.191	
Only injection	0.555** (.208)	1.743**	-0.555** (.208)	0.574**	
Pills and injection	0.489** (.157)	1.631**	-0.489** (.157)	0.613**	
Optimistic about future					
Disagree	-0.024 (.166)	0.976	0.024 (.166)	1.024	0.034
Neutral	-0.213 (.122)	0.808	0.213 (.122)	1.237	
Agree	0.237* (.118)	1.268*	-0.237* (.118)	0.789*	
Class Share	0.559		0.441		

a. Wald tests for equal preference estimates and equal variances across classes.
b. The coefficients are log odds ratios with effect coding.
*p<0.05 **p<0.01 ***p<0.001

Table 3-6 Model statistics of mixed logit, 6-class latent class logit, 5-class 2-scale-class latent class logit, and 2-class random effect latent class logit [a]

	LL	BIC	Npar	df	Entropy R^2	Pred. Err.
MXL	-4218.12	8587.38	24	519	1	0.130
6-class LCL	-3976.97	8438.82	77	466	0.832	0.158
5-class 2-sclass SLCL	-3993.79	8403.18	66	477	0.773	0.157
2-class RELCL	-4021.04	8350.64	49	494	0.681	0.116

a. None of the models include individual characteristics.

MXL: mixed logit; LCL: latent class logit; SLCL: scale-adjusted latent class logit; RELCL: random effect latent class logit; LL: log-likelihood; BIC: Bayesian Information Criteria; Npar: number of parameters; df: degree of freedom; Pred. Err.: prediction error.

Table 3-7 Cross-tab between class membership in the 5-class scale-adjusted latent class logit model and class membership in the 2-class random effect latent class model [a]

		RELCL		
		Class 1 (cost-effective)	**Class 2 (side-effect focused)**	**Total**
SLCL	**Class 1 (cost dominant)**	102 (91.1%)	10 (8.9%)	112
	Class 2 (treatment burden dominant)	20 (15.9%)	106 (84.1%)	126
	Class 3 (nausea dominant)	25 (19.1%)	107 (80.9%)	131
	Class 4 (effectiveness dominant)	91 (95.8%)	4 (4.2%)	95
	Class 5 (effect stability dominant)	58 (75.3%)	19 (24.7%)	77
	Total	296	245	541

a. Random effect latent class model includes individual characteristics.
SLCL: scale-adjust latent class logit; RELCL: random effect latent class logit.

Chapter 4 Using Latent Class Analysis To Analyze Preference Heterogeneity In Best-Worst Scaling Data: A Comparison of Latent Class Logit And Standard Latent Class Models

Mo Zhou, MPA, MHS; Karen Bandeen-Roche, PhD; John FP Bridges, PhD

Abstract

Latent class logit (LCL) models have been commonly used to analyze preference heterogeneity in stated-preference studies. With an increasing use of best-worst scaling (BWS) to assess priorities in healthcare in the past few years, researchers started to apply LCL in BWS data. However, information criteria often fail to identify the best-fitted model when LCL is used to segment discrete choice data from BWS. This paper proposes alternative models for this emerging preference elicitation technique, specifically, applying the standard LC model on the best or worst choices or best-worst scores.

We test the model using data from a national survey among patients with type 2 diabetes on barriers and facilitators for diabetes self-management. The survey contains a BWS (case 1) experiment with 11 choice tasks generated from a balanced-incomplete-block design. The standard LC model with best and worst choices both identifies five classes based on minimized BIC. The LC model with best-worst scores includes four classes, but the segments are comparable to those from the best/worst choice model. Instead of LCL, researchers should consider more flexible model specifications to improve model fit when analyzing BWS data.

4.1 Introduction

Best-worst scaling (BWS) has been increasingly used to measure priorities in health (Cheung et al., 2016). Unlike the traditional ranking and rating methods, BWS repeatedly presents subsets of objects to be prioritized and asks respondents to select the most and least important object in each set (Muhlbacher et al., 2016). Based on repeated comparison and choices, BWS ranks all objects in terms of importance. In healthcare, the priorities of patients and other stakeholders measured by BWS have been used to direct clinical care and policy (Cheung et al., 2016). For example, knowledge about patients' priorities on clinical outcomes allows clinicians to tailor treatment plan to patients' goals (Lynd et al., 2016).

Recent movements towards patient-centered drug development from the regulatory agency (Califf, 2017) and patient-centered care among healthcare providers (Epstein et al., 2011) emphasize the importance of learning heterogeneity in patient preferences and priorities. Latent class analysis (LCA) is a common method to study unobserved preference heterogeneity. It groups individuals into like-minded clusters based on their preference or priority (Cunningham et al., 2008; Hole, 2008). Despite its wide application in stated-preference studies, LCA has not been commonly used to analyze heterogeneity in priorities, especially those measured by BWS.

A systematic review we conducted earlier found that only five articles applied segmentation methods to study heterogeneity in priority measured by BWS case 1 (object case). Four studies used latent class logit (LCL) model (Feudtner et al., 2015; Fraenkel et al., 2015; Virudachalam et al., 2016; Yan et al., 2015), and one study used k-mean clustering analysis for segmentation (Liang et al., 2016). None of the studies explained model specification or model selection process. Given that the application of LCA in BWS data has not been fully examined, this study sought to show the challenges of using LCL, the most commonly used segmentation method in stated-preference studies, to analyze heterogeneity in BWS data and propose alternative model specifications. We hypothesize that more flexible model specification will improve model fit, leading to more reliable preference estimates and prediction of choices to support policy and clinical decision-making.

This study uses BWS data from a national survey among patients with type 2 diabetes on the barriers and facilitators for diabetes self-management. Type 2 diabetes is chosen as a case study because it represents a chronic illness that requires meaningful patient engagement to improve disease management and outcomes. Learning how the barriers and facilitators for diabetes self-management vary among different patients will enable clinicians to tailor treatment plan based on patient priority and improve treatment adherence as well as outcomes. Given that diabetes affects nearly 30 million individuals in the US (CDC, 2014), the result will have significant clinical implications. The remainder of this paper is organized as follows. Section 2 describes the BWS survey and sample characteristics. Section 3 shows the challenge of using LCL model in BWS data and proposes alternative models to analyze heterogeneity in priority. Section 4 presents the results and section 5 discusses the methodological and clinical implications.

4.2 Data

We collected data on barriers and facilitators for diabetes self-management among patients with type 2 diabetes using BWS case 1 (object case). 11 factors were purposely selected from literature reviews, focus group interviews, and repeated feedback from local patients to represent both known barriers and facilitators of patients' diabetes management (Table 4-1). A balanced-incomplete block design (BIBD) generated 11 choice tasks. Each task included a subset of five factors and asked respondents to choose the factors with the best impact and worst impact on their own diabetes self-management. Each factor appeared five times and was compared to every other factor twice across the 11 choice tasks.

[TABLE 4-1 INSERT HERE]

Figure 4-1 presents one example choice task. Respondents were asked to complete all 11 choice tasks that were presented in random order. Detailed explanation of the factors was provided to respondents before they answered the choice tasks. The survey questions as well as the framing of the information was pretested and piloted among local patients to make sure they were understandable. In addition to BWS choice tasks, the survey also included questions on the respondents' diabetes history, current disease management, health status, and personalities using Likert scales.

[FIGURE 4-1 INSERT HERE]

A national survey was conducted among patients with type 2 diabetes through the GfK KnowledgePanel in 2015. It is an online panel that provides sampling coverage of 97% of the US adult population (Couper, 2000), and has been shown to have over 4,000 potential patients with type 2 diabetes, with a high proportion of complex patients (Safford et al, 2007). 554 respondents filled out the survey. 536 who completed all choice tasks and did not select the same object as both best and worst in any of the choice tasks are included in the analysis.

The mean age was 61.6 years. 46.8% of the respondents were female. African Americans and Hispanics were oversampled (22.6% and 21.6%, respectively) due to higher prevalence of diabetes (CDC, 2014) and lower adherence rates to treatments (Egede et al, 2011). Over 90% of the sample had at least high school degree, and almost half had above $50,000 annual income. The average length of diabetes diagnosis was 11 years. About 90% of sample measured hemoglobin A1c at least once in the past six months. 39.9% had A1c below 7 percent, and 58.8% had no complications. Only 8.2% of the sample was not taking any diabetes treatments. Majority (73.9%) reported good, very good, or excellent overall health. Half of the respondents reported having self-control and only 16.6% reported to be risk-taker. (Table 4-2) GfK calculated sample weights to rebalance the sample to national norms for the assessment of aggregate results.

[TABLE 4-2 INSERT HERE]

4.3 Model specifications

4.3.1 *Latent class logit model*

We first estimated a latent class variant of the sequential best-worst model (Marley and Louviere, 2005) to analyze heterogeneity in priority given its common application in discrete choice data as well as in the few BWS studies in health. The model assumes that individuals choose the best and worst options in a choice task in two independent steps. They first choose the best object with the highest utility and then the worst object with the lowest utility among the remaining options. The utility an individual *n* derives from factor *i* in choice task *t* is:

$$U_{nit} = \beta_{ni} x_{nit} + \varepsilon_{nit} \qquad [1]$$

where x_{nit} is usually the characteristics for object i, but in the case of BWS (object case), it is an object specific dummy variable for factor i, and β_{ni} then measures the utility individual n assigns to object i. ε_{nit} is a random disturbance term that represents the random fluctuations when the decision maker makes choices (McFadden, 1974). When ε is assumed to follow an independent type I extreme value distribution, the probability of individual n choosing object i as the best in choice task t becomes:

$$P_{nt}(i) = \frac{\exp(\beta_{ni} x_{nit})}{\sum_{j=1}^{5} \exp(\beta_{nj} x_{njt})} = \frac{\exp(\beta_{ni})}{\sum_{j=1}^{5} \exp(\beta_{nj})} \qquad [2]$$

given that xs are object specific dummies. To achieve identification, one utility parameter needs to be normalized to zero. This study uses effect coding and therefore β_i measures the relative utility for object i compared to the mean utility across all objects. The probability of individual n choosing object i as best and k as worst in choice task t is:

$$P_{nt}(i,k) = \frac{\exp(\beta_{ni})}{\sum_{j=1}^{5} \exp(\beta_{nj})} \times \frac{\exp(-\beta_{nk})}{\sum_{j \neq i} \exp(-\beta_{nj})} \qquad [3]$$

assuming the probability of an object being chosen as worst depends on the magnitude of the negative relative utilities of all objects in the remaining choice set. That is, the object with the largest negative relative utility (i.e. the smallest positive relative utility) will have the highest probability to be chosen as worst.

LCL model assumes the individual utility parameters to be random draws on a discrete distribution with distinct groups (McFadden, 1986). With different class-specific parameters and class membership probability, LCL captures the inter-personal preference heterogeneity as well as the intra-personal correlation in responses to different choice tasks. Suppose there are Q distinct groups, the likelihood of a sequence of choices from individual n is:

$$L_n(y) = \sum_{q=1}^{Q} \pi_{q|n} \prod_{t=1}^{11} P_{nt}\left(y_t \middle| \beta^q\right) \qquad [4]$$

where y is a vector of best and worst choices from the 11 choice tasks, y_t is the best and worst choices from choice task t, $\pi_{q|n}$ is the probability of person n belonging to class q, and β^q is a vector of class-specific utility parameters for the factors in class q. We estimated the LCL model with one to ten classes with effects coding. Bayesian Information Criterion (BIC) was used to identify the appropriate number of classes, given that it has been shown to have better performance than other criteria (Nylund et al, 2007). Each model was replicated 10 times with random starting seeds to look for global maxima.

Figure 4-2 shows the BICs from the LCL models. As the number of classes increases from one to ten, BIC continues decreasing, indicating model over-fit. For a sensitivity check, we 1) allowed different scale for the worst choices in LCL, 2) allowed two scale classes to account for different consistency of choices among different respondents, and 3) estimated the best only and worst only models to avoid the dependency between best and worst choices. However, none of the adjustments led to a best-fitted model with fewer than ten classes. Instead of assuming sequential best-worst, we also considered alternative choice behavior assumption. The maximum difference logit (max-diff) model assumes that individuals consider all possible pairs within a choice set and choose the pair with maximum difference in utility (Marley and Louviere, 2005; Marley et al., 2008). However, estimates from the max-diff model were almost identical to those from LCL, which is consistent with findings from previous studies (Flynn et al., 2008). Alternative model specifications are needed to analyze heterogeneity in BWS data, and we propose two models in this study.

[FIGURE 4-2 INSERT HERE]

4.3.2 *Standard latent class model on best and worst choices*

We first apply a standard LC model on the 11 choice tasks as 11 categorical outcome indicators. We start with assuming choices across the tasks are independent given the class-specific preferences. However, the best and worst choices within a choice task are dependent. Allowing dependency between best and worst across 11 tasks will dramatically complicate the model and reduce degree of freedom. Given that we have relatively limited sample size for 11 outcome indicators, the best and worst choices are analyzed separately. We assume the worst choice is selected out of all five options in a choice task, instead of five options less the best as suggested in the sequential best-worst model, to ensure same choice set across respondents. Given that there could be a mixture of choice behaviors among respondents (i.e. best-worst, worst-best, or simultaneous choices) that is unknown, this assumption is a valid compromise. Each indicator then has five possible outcomes. Using the notation above, the probability of individual n choosing object i as best (or worst) in choice task t can be written as:

$$P_{nt}(i) = \frac{\exp(\beta_{nit})}{\sum_{j=1}^{5} \exp(\beta_{njt})} \qquad [5]$$

where β_{nit} is individual n's utility (or disutility) for object i in choice task t. When there are Q distinct classes, the utility for object i in choice task t of respondents in class q is specified as $\beta_{nit} = \beta_{it} + \beta_{it}^{q}$. The likelihood of a sequence of choices from individual n is the same as equation [4].

Compared to LCL, the standard LC model allows the utility parameter (*β*) for each object to vary across different choice tasks. Restricted by a cross-task constant utility constraint, the utility parameter in LCL measures the average utility for an object across all choice tasks. With 11 choice tasks and potential intra-personal inconsistency in choices compared to preferences, there can be a wide variety of choice sequences, leading to hundreds of utility values for each object across respondents. Segmentation on 11 indicators with hundreds of potential outcomes can lead to many classes. The standard LC model relaxes such constraint. Because there are only five possible outcomes in each choice task, the utility parameters take on a much smaller set of values. We hypothesize that despite a larger number of parameters, the standard LCL model leads to more parsimonious segmentation results.

We estimate the standard LC model on best and worst choices separately with one to ten classes and use BIC to choose the most appropriate number of classes. Each model is replicated 10 times with random starting seeds to look for global maxima. As the choice tasks repeatedly present subsets of objects, it is likely that choices from some tasks are correlated. Based on the output from the independent model, we relax the local independency assumption by allowing association between indicators with large bivariate residuals. Suppose choice task t_1 is correlated with choice task t_2. The two dependent variables then serve as a joint dependent indicator. The likelihood of a sequence of choices becomes:

$$L_n(y) = \sum_{q=1}^{Q} \pi_{q|n} P_{nt_1t_2}\left(t_1 = i, t_2 = j \middle| \beta^q\right) \prod_{t \notin t_1,t_2} P_{nt}\left(y_t \middle| \beta^q\right) \qquad [6]$$

where the linear term in the logit model for $P_{nt_1t_2}\left(t_1 = i, t_2 = j \middle| \beta^q\right)$ is

$$\beta_{nijt_1t_2} = \beta_{it_1} + \beta_{it_1}^q + \beta_{it_2} + \beta_{it_2}^q + \beta_{ijt_1t_2}.$$

We then compare the estimates from models with and without local independence assumption.

Using the estimates from the best-fitted model, we calculate the average conditional probability for each object across the five choice tasks in which it appears. Because of the balanced design in the survey, the average conditional probabilities are comparable across objects and therefore are used to generate a full ranking of the 11 objects for each class. We then perform a multinomial logistic regression to examine the individual characteristics that are associated with class membership. Sampling weights are taken into account when calculating standard errors, the latent class distribution, and covariate effects.

4.3.3 *Standard latent class model on best-worst scores*

One weakness of the standard LC model above is that the segmentation is based on only best or worst choices. In order to cluster respondents based on both choices without considering specific choice behavior, we apply a standard LC model on best-worst (BW) scores. BW score of each object is calculated for each respondent by subtracting the number of times an object is selected as worst from the number of times the object is selected as best. Given that each object appears five times in the 11 choice tasks, BW score ranges from -5 (i.e., an object is selected as worst every time it appears) to 5 (i.e., an object is selected as best every time it appears). A standard LC model is applied to the 11 BW scores as 11 outcome indicators. Again, we first assume local independence between the BW scores for different objects. To reduce the dimension of LC model and avoid over-fitting, BW scores are grouped into five categories, i.e., -5 to -

4, -3 to -1, 0, 1 to 3, and 4 to 5. The probability of an object *i* getting a BW score that is in category *m* from individual *n* in class *q* is:

$$P_{ni}\left(m|\beta^q\right)=\frac{\exp\left(\beta_{mi}+\beta^q_{mi}\right)}{\sum_{m'=1}^{5}\exp\left(\beta_{m'i}+\beta^q_{m'i}\right)} \qquad [7]$$

where β_{mi} is the mean log odds ratio for object *i* to have a BW score in category *m*, and β^q_{mi} is the difference in mean log odds ratio for object *i* to have a BW score in category *m* between all respondents and those in class *q*. Because the 5-category BW score is an ordinal indicator, we restrict that $\beta^q_{mi}=\beta^q_i \cdot z_{mi}$, where z_{mi} is the score assigned to category *m* of BW score for object *i*, which yields an ordinal logit model. The likelihood of a sequence of BW scores from individual *n* is:

$$L_n\left(y\right)=\sum_{q=1}^{Q}\pi_{q|n}\prod_{i=1}^{11}P_{ni}\left(m|\beta^q\right) \qquad [8]$$

where *y* is a vector of BW scores for the 11 objects. Intuitively, LC is performed on 11 ordered logit regressions of BW scores. We estimate the model with one to ten classes and use BIC to choose the appropriate model. Based on the bivariate residuals between indicators in the model output, we relax the local independence assumption by allowing correlation between indicators, and then compare the estimates to those from the restricted model. Multinomial logit regression is used to explore the correlation between individual characteristics and class membership in the best-fitted model. Standard errors, latent class distribution, and covariate effects are adjusted using sampling weights.

4.4 Results

All factors were selected at least 10 times as best and at least 22 times as worst in each choice task across all respondents, indicating large variation in choices between respondents. By allowing task-specific parameters, the standard LC model on both best/worst choices and BW scores dramatically reduced the number of classes and identified four to five classes based on minimized BIC. Large bivariate residuals exist between outcome indicators in both best/worst choice model and BW score model, suggesting correlation between choice tasks and between BW scores of different factors. However, relaxing local independence assumption did not significantly change the segmentation results.

4.4.1 *Standard latent class model on best and worst choices*

The optimal LC model on best choices identified five facilitator classes. Healthcare providers, support from family and friends, access to healthy food, staying motivated, and the ability to pay were the greatest facilitator for diabetes self-management among respondents in class 1 (provider-driven class, 26.2% of the sample), class 2 (support-driven class, 21.9%), class 3 (healthy food-driven class, 21.0%), class 4 (motivation-driven class, 18.1%), and class 5 (affordability-driven class, 12.8%) respectively (Figure 4-3). Respondents' knowledge about diabetes was the second most influential facilitator in all classes except class 2 where family commitment was also important.

[FIGURE 4-3 INSERT HERE]

In terms of class characteristics, a 10-year increase in age was associated with a 5% increase (p=0.003) in the probability of being in the provider-driven class (Table 4-3). African Americans were 12.7% (p=0.024) more likely to be in the affordability-driven class and 9.1% (p=0.049) less likely to be in the health food-driven class compared to the white counterparts. Hispanics were 17.0% (p=0.002) less likely to be in the provider-driven class than whites. Patients with high school and bachelor degrees were 13.0% (p=0.024) and 15.1% (p=0.037), respectively, more likely to be in the support-driven class than those without high school degree. Poorer overall self-reported health status was correlated with higher probability of being in the support-driven class ($p<0.05$), while patients who reported poor health were 24.0% (p=0.050) less likely to be in the provider-driven class. In terms of personalities, individuals who reported having self-control were 11.8% more likely to be in the motivation-driven class than the rest (p=0.024). Patients who reported to be risk-takers were 15.5% (p=0.031) more likely to be in the support-driven class and 17.4% (p=0.001) less likely to be in the provider-driven class.

[TABLE 4-3 INSERT HERE]

The standard LC model with the worst choices identified five barrier classes. Work commitments and local events were the greatest barriers for diabetes self-management for respondents in class 1 (work-committed class, 38.4% of the sample). Ability to pay was the single leading barrier for class 2 (affordability-lacking class, 22.2%). Staying motivated was the primary barrier for class 3 (motivation-lacking class, 19.2%), followed by other health conditions. Support from family and friends and staying motivated were the most important barriers for class 4 (support-lacking class, 10.6%). Access to healthy food was the most influential barriers for class 5 (healthy food-lacking class, 9.7%), while lack of motivation, ability to pay, and other health conditions were also important (Figure 4-4).

[FIGURE 4-4 INSERT HERE]

Hispanics were 6.6% (p=0.004) less likely to be in the healthy food-lacking class, while Asians were 19.3% ($p<0.001$) less likely to be in the affordability-lacking class than the white counterparts (Table 4-4). Patients with high school and bachelor degrees were 13.3% ($p=0.041$) and 16.9% ($p=0.025$), respectively, more likely to be in the motivation-lacking class than those without high school degrees. Income was positively correlated with the probability of being in the motivation-lacking class ($p<0.05$). Poorer health status was associated with lower probability of being in the healthy-food-lacking class and higher probability of being in the motivation-lacking class ($p<0.05$). Patients with hemoglobin A1c below 7% were 5.6% ($p=0.022$) less likely to be in the healthy food-lacking class. Patients who didn't know their A1c level were 12.4% ($p=0.034$) more likely to be in the health food-lacking class. Individuals who had self-control were 16.0% more likely to be in the work-committed class ($p=0.017$). Risk-takers were 14.1% less likely to be in the motivation-lacking class ($p=0.006$).

[TABLE 4-4 INSERT HERE]

4.4.2 *Standard latent class model on best-worst scores*

The 5-class LC model with BW scores has the lowest BIC (15751.99), but because one class in the model only has 3% of the respondents, we selected a more parsimonious 4-class model (BIC=15754.21), which also has the lowest consistent AIC among all. Figure 4-5 shows the conditional probability of each object getting a BW score in each category by class. In class 1 (27.9% of the sample), knowledge and motivation were the greatest facilitators for diabetes self-management, while ability to pay was the leading barrier. Class 2 (24.5%) identified support from others, healthcare providers, and family commitment as the main facilitators and other health condition and staying motivated the major barriers. Knowledge, motivation, healthcare providers, and access to healthy food were all facilitators for respondents in class 3 (24.0%), while local events and work commitments were the main barriers. Class 4 (23.7%) chose knowledge and ability to pay as important facilitators, followed by access to healthy food and healthcare providers, and selected support from others, motivation, other health conditions, and family commitment as major barriers.

[FIGURE 4-5 INSERT HERE]

In terms of class characteristics, individuals who were widowed, divorced or separated and those who were never married were 13.8% ($p=0.005$) and 22.7% ($p<0.001$), respectively, less likely to be in class 2 compared to those who were married. Patients with high school and bachelor degrees were 23.4% and 21.9%, respectively, more likely to be in class 2 ($p<0.001$), while 24.1% ($p=0.037$) and 27.4% ($p=0.034$), respectively, less likely to be in class 1 than those without high school degree. Patients with over $75,000 annual income were 16.2% more likely to be in class 4 than those with under $25,000 annual income ($p=0.040$). Poorer overall health status was correlated with higher probability of being in class 2 ($p<0.05$). Patients with hemoglobin A1c below 7% were 10.4% less likely to be in class 1 ($p=0.032$). Individuals who reported having self-control were 16.7% ($p=0.002$) more likely to be in class 3 and 13.9% ($p=0.036$) less likely to be in class 4. (Table 4-5)

[TABLE 4-5 INSERT HERE]

Table 4-6 describes the cross-tab between the class membership from BW score model and the class membership from best/worst choice model. It summarizes how the BW score model combines and redistributes the barrier and facilitator classes. Class 1 mainly took respondents in the affordability-lacking class (71.8%), with half of the motivation-driven class who may be the same group of individuals. Class 2 was mainly from the support-driven class, with half of the motivation-lacking class that may be related. Majority of class 3 were from the work-committed class. Class 4 combined affordability-driven class with motivation-lacking and support-lacking classes. Segmentation results from the BW score model were consistent with those from the best/worst choice models. Performing LC on BW score just incorporates the facilitator and barrier classes generated from the best and worst choices with consideration of the correlation between the two parts.

[TABLE 4-6 INSERT HERE]

4.5 Discussion

This paper identified the issue with applying LCL model in BWS object case data to analyze heterogeneity, specifically, the challenge of finding the appropriate number of classes using information criteria such as BIC. The issue rises when there are many choice tasks in a survey and there is large variation in the choice sequence among respondents. This is because in these cases the cross-task equal utility constraint in the LCL model leads to large variation in the utility parameters across respondents. Segmentation on a large set of indicators, each with hundreds of possible outcomes, can lead to a large number of classes. The BWS survey in this study had 11 prioritization factors and choice tasks as well as large variation in choice sequence among the respondents. BIC from the LCL model was therefore unable to reach minimum within ten classes.

LCL model may lead to parsimonious segmentation results only when there are a limited number of choice patterns across respondents, meaning that there are only a few priority types among respondents and there is little inconsistency between choices and priorities (i.e., disturbance terms in the utility function barely affect choices). Given that over half of BWS case 1 studies in health include over 10 prioritization objects (Cheung et al., 2016), LCL model is unlikely to lead to practical segmentation results. Increasing sample size may also exacerbate the issue due to additional priority types and larger inconsistency in choices among additional respondents, especially when we study the priority of a diverse population such as the diabetes patients in this paper. The few studies in the current literature that used LCL to analyze BWS data selected the number of classes based on interpretation of the classes. Model fit, however, should be considered to ensure reliable estimates and predictions.

We proposed two alternative models to analyze heterogeneity in BWS data in this paper. We first used a standard LC model to separately analyze best and worst choices, allowing utility parameter for each object to vary in different choice tasks. This is equivalent to a LC analysis on 11 multinomial logistic regressions with only object specific constant as predictor. Because respondents make choices based on a comparison of options within a choice task and their relative utility, the task-specific utility assumption is consistent with choice behavior. Another assumption in the model is that both best and worst choices are selected from all options in a choice set. Given that sequential best-worst and max-diff assumptions often lead to similar results (Flynn et al., 2008), we argue that the assumption on choice behaviors would not affect segmentation results. However, future research can test this hypothesis by allowing sequential best-worst, worst-best, and max-diff and examining the impact of choice behavior assumption on segmentation results. Overall, the flexible model specification led to more parsimonious segments, despite additional parameters to be estimated.

One weakness of the standard LC model with choices as outcome is that it is difficult to accommodate both best and worst choices in the model when there are many choice tasks (i.e. outcome indicators). This is because the dependence between the best and worst choices requires additional correlation parameters that further complicate the model structure. To address this issue and incorporate both best and worst choices without further reducing degree of freedom, we applied a standard LC model on BW scores, a single measure that reflects both choices. The model led to parsimonious segments that are comparable to those from the best/worst choice models. Specifically, segmentation based on BW scores combines and regroups classes from the individual best and worst choice models. However, future research can consider including both best and worst choices in the choice model and allowing local dependence between the two outcomes when there are fewer prioritization objects and larger sample size. The results can be compared to the BW score model.

We found substantial heterogeneity in both barriers and facilitators for diabetes self-management among patients with type 2 diabetes. Older patients were more likely to identify healthcare providers as the main facilitator for diabetes self-management, probably because of more access to healthcare due to Medicare. African Americans were not likely to benefit from access to healthy food, but were likely to report ability to pay as a facilitator for their diabetes management. Hispanics did not have an issue with access to healthy food, but were unlikely to receive help from healthcare providers. Support from family and friends and family commitment facilitated diabetes management among individuals who were married. The more educated patients also benefited the most from support from others. They did not have an issue with ability to pay, but could have trouble staying motivated. Due to sufficient financial resources, the high-income group had good access to healthcare providers and healthy food. They reported their ability to pay facilitated their diabetes management, but like the educated group, they could have trouble staying motivated. In addition, family commitment and lack of support from others could also negatively affect their disease management.

Patients with poorer overall health were more likely to report other health conditions jeopardized their diabetes management. They were likely to lack motivation, but benefit most from support from others. Healthcare providers were unlikely to be a major facilitator for those with the poorest health, possibly due to limited healthcare access. Patients with hemoglobin A1c above 7% were more likely to report ability to pay as a barrier and rely on knowledge and motivation to manage diabetes. Work commitment was the leading barrier for patients who reported having self-control, while motivation was the major facilitator in this group. Individuals who reported to be risk-takers did not have issue staying motivated. Their disease management was not likely to be facilitated by healthcare providers but support from others.

Majority of the studies in the currently literature focus on examining the impact of individual factors, such as economic resources, social networks, and health knowledge, on diabetes management, and few studies ask patients to prioritize them. This study fills the knowledge gap by ranking the barriers and facilitators from the patients' perspective and, more importantly, segmenting patients based on their priorities to learn the heterogeneity among different individuals. Segmentation based on priorities sometimes gives contradicting findings from stratification on individual socio demographic characteristics. For example, Chelbowy et al. (2013) found gender differences in the barriers and facilitators for diabetes self-management among African American adults, which was not observed in our study. This is possibly because other individual characteristics (e.g., income, education, personality) were controlled for in the multivariate analysis of the correlation between an individual covariate and class membership. By controlling for potential confounding, we were able to identify the key factors associated with different priority types.

Our findings are consistent with the current literature on that the variation in diabetes self-management as well as the barriers and facilitators reflects individuals' knowledge, option, and personality more than patients' age, gender, or culture (Onwudiwe et al., 2011). We found that access to healthcare, social support and motivation were the three main themes affecting diabetes self-management. According to our findings, expanding healthcare coverage for the uninsured or under-insured, motivating those who are educated, high-income or sick, reducing work pressure for the motivators, and enhancing social support for the high-incomes would be the most effect way to improve diabetes management and health outcome in these groups. Clinicians can consider such diversity and tailor their treatment and disease management plans based on individual needs.

There are a few limitations in this study. First, the sample size is relatively small given the number of objects examined in the study. As a result, we were not able to allow correlation between all pairs of choice tasks with large bivariate residuals. However, the local dependence constraint was relaxed for the strongest correlated tasks and it did not significantly change the segmentation results. Future research with larger sample size can allow more correlated pairs or even incorporate both best and worst choices in the model.

Moreover, we conducted a post-estimation analysis to examine class characteristics to reduce the complexity of the model. With larger sample size, future research can consider including individual characteristics in the multinomial logit function for class membership to improve segmentation.

Second, patient insurance information is missing in the data. As we included most socio demographic characteristics, health and disease-related variables, and personalities traits, we were unable to control for health insurance status. Given some findings such as the insignificant correlation between income and the probability of patients reporting ability to pay as a main barrier, we hypothesize that health insurance status plays an important role in diabetes self-management. If so, the results are confounded. In addition, the limited employment information (i.e., employed, unemployed, retired) did not allow us to identify the specific group who reported work commitment as the leading barrier. It may be valuable to further examine these two factors in future research.

Third, respondents in the sample have higher income and education level than the general population, even among the minorities. This could be the reason why we found that African Americans were more likely to report ability to pay as a facilitator for diabetes self-management. Although we used sample weights to adjust standard errors, latent class distribution, and covariate effects and controlled for socioeconomic status such as income and education, the priorities from the low-income, low-education minority groups could still be underrepresented in the sample. Therefore, researchers should use caution to generalize the results.

Finally, we were not able to validate the individual priority as well as the priority types revealed in our analysis with those obtained from other sources. We compared the priority segments across different model specifications and tested the convergent validity. However, it is unclear if the BWS survey revealed the true priorities among respondents. Future research can use other prioritization methods, such as ranking, in addition to BWS to study priority to allow validation across methods.

Despite the limitations, this paper identifies the issue of applying LCL model to analyze BWS data and proposes alternative model specifications that are more suitable for the BWS data structure. Depending on the study objective, researchers can apply the standard LC model on either best/worst choices to separately examine barriers and facilitators or BW score to capture both answers. In either model, researchers should consider flexible model specification to allow conditional probability for each object to vary in different choice scenarios. Improved model fit would lead to reliable estimates and predictions to support decision-making.

Reference

Califf RM. Benefit-risk assessments at the US Food and Drug Administration: finding the balance. JAMA. 2017. doi: 10.1001/jama.2017.0410. [Epub ahead of print]

Cheung KL, Wijnen BF, Hollin IL, et al. Using best-worst scaling to investigate preferences in health care. Pharmacoeconomics. 2016;34(12):1195-1209.

Chlebowy DO, Hood S, and LaJoie AS. Gender differences in diabetes self-management among African American adults. Western Journal of Nursing Research. 2013;35(6):703-21.

Couper MP. Review: web surveys: a review of issues and approaches. The Public Opinion Quarterly. 2000;64(4):464-94.

Cunningham CE, Deal K, Rimas H, et al. Using conjoint analysis to model the preferences of different patient segments for attributes of patient-centered care. The Patient: Patient-Centered Outcomes Research. 2008;1(4):317-30.

Epstein RM and Street RL. The values and value of patient-centered care. Annals Family Medicine. 2011;9(2):100-3.

Flynn T, Louviere J, Peters T, et al. Estimating preferences for a dermatology consultation using best-worst scaling: comparison of various methods of analysis. BMC Medical Research Methodology. 2008;8:76.

Hole AR. Modelling heterogeneity in patients' preferences for the attributes of a general practitioner appointment. Journal of Health Economics. 2008;27(4):1078-94.

Lynd LD, Traboulsee A, Marra CA, et al. Quantitative analysis of multiple sclerosis patients' preferences for drug development: a best-worst scaling study. Therapeutic Advances in Neurological Disorders. 2016;9(4):287-96.

Marley A, and Louviere J. Some probabilistic models of best, worst, and best-worst choices. Journal of Mathematical Psychology. 2005;49:464-80.

McFadden D. Conditional logit analysis of qualitative choice behavior. In: P. Zarembka, editor. Frontiers in econometrics. New York: Academic Press; 1974.

McFadden D. The choice theory approach to market research. Marketing Science. 1986;5(4):275-97.

Muhlbacher AC, Kaczynski A, Zweifel P, et al. Experimental measurement of preferences in health and healthcare using best-worst scaling: an overview. Health Econ Rev. 2016;6(1):2.

Nylund KL, Asparouhov T, and Muthén BO. Deciding on the number of classes in latent class analysis and growth mixture modeling: a Monte Carlo simulation study. Structural Equation Modeling. 2007;14:535–569.

Onwudiwe NC, Mullins CD, Winston RA, et al. Barriers to self-management of diabetes: a qualitative study among low-income minority diabetics. Ethnicity & Disease. 2011;21(1): 27-32.

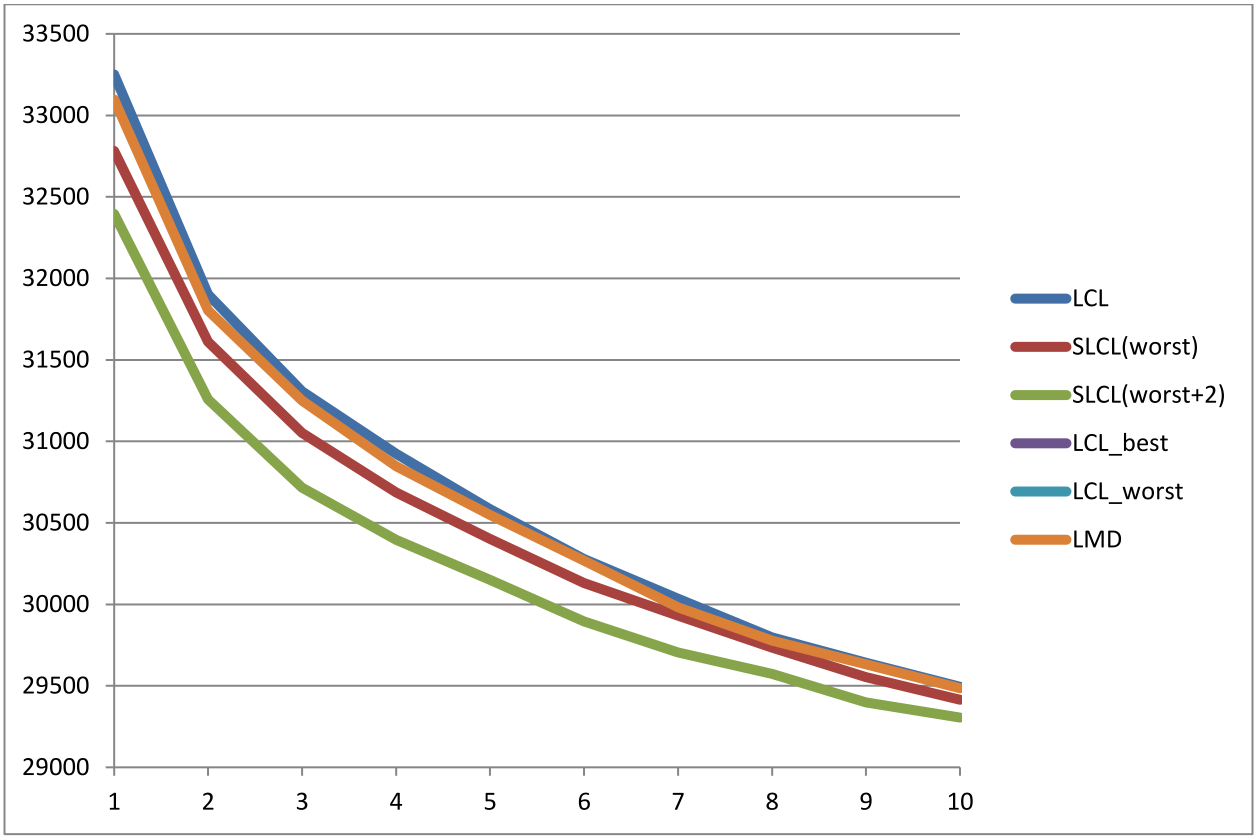
Safford MM, Allison JJ, Kiefe CI. Patient complexity: More than comorbidity. The vector model of complexity. Journal of General Internal Medicine. 2007;22:382-90.

Figure 4-1 Sample choice task in the best-worst scaling survey

Consider the following things that can have a positive or negative impact on your own diabetes self-management. Which of the following things is the **best** and which is the **worst** in terms of impact on your own diabetes self-management?

Things impacting your own diabetes self-management	Best (pick only one)	Worst (pick only one)
Access to healthy food	☐	☐
Healthcare providers	☐	☐
My ability to pay	☐	☐
Local events	☐	☐
Family commitments	☐	☐

Figure 4-2 Bayesian information criterion (BIC) from the latent class logit models with one to ten classes



LCL: latent class logit;
SLCL (worst): scale-adjusted latent class logit with different scale for worst choices;
SLCL (worst+2): scale-adjusted latent class logit with different scale for worst choices and two scale classes for choice consistency;
LCL_best: latent class logit with best choices;
LCL_worst: latent class logit with worst choices;
LMD: latent class logit with max-diff

Figure 4-3 Average conditional probability of each factor being chosen as best in a choice task by facilitator class from the standard latent class model

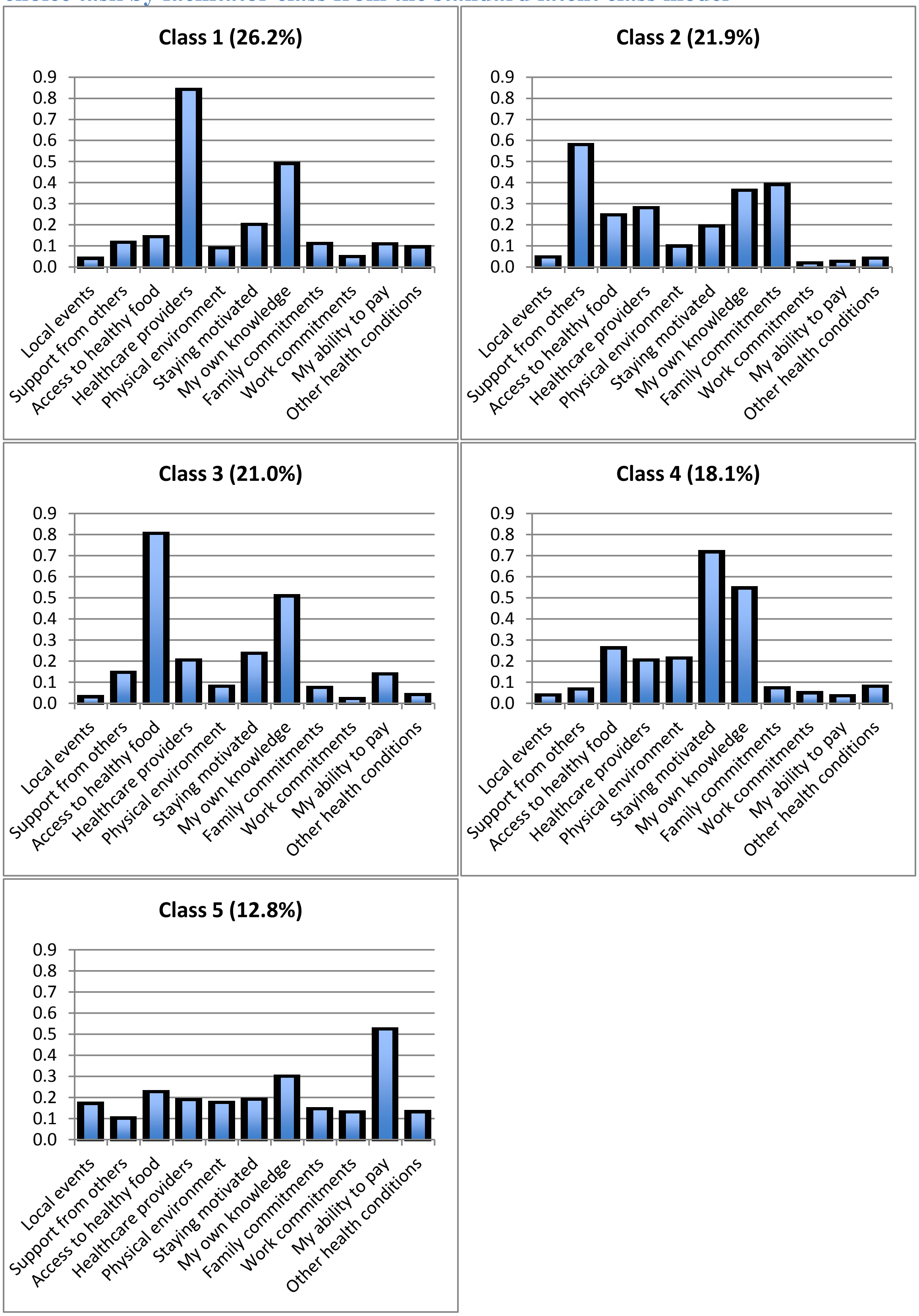


Figure 4-4 Average conditional probability of each factor being chosen as worst in a choice task by barrier class from the standard latent class model

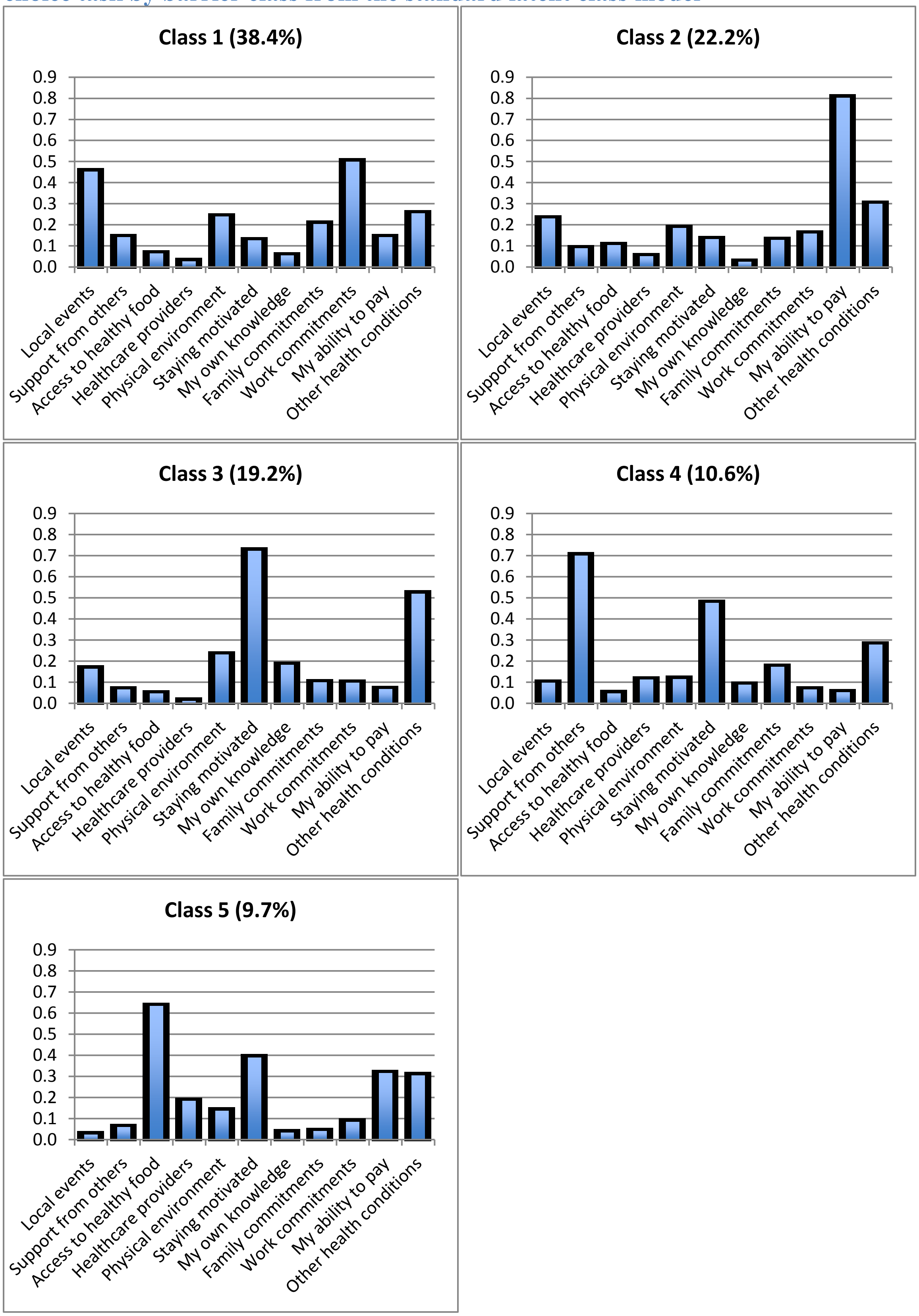


Figure 4-5 Conditional probability of the 5-category best-worst score for each object by class from the standard latent class model

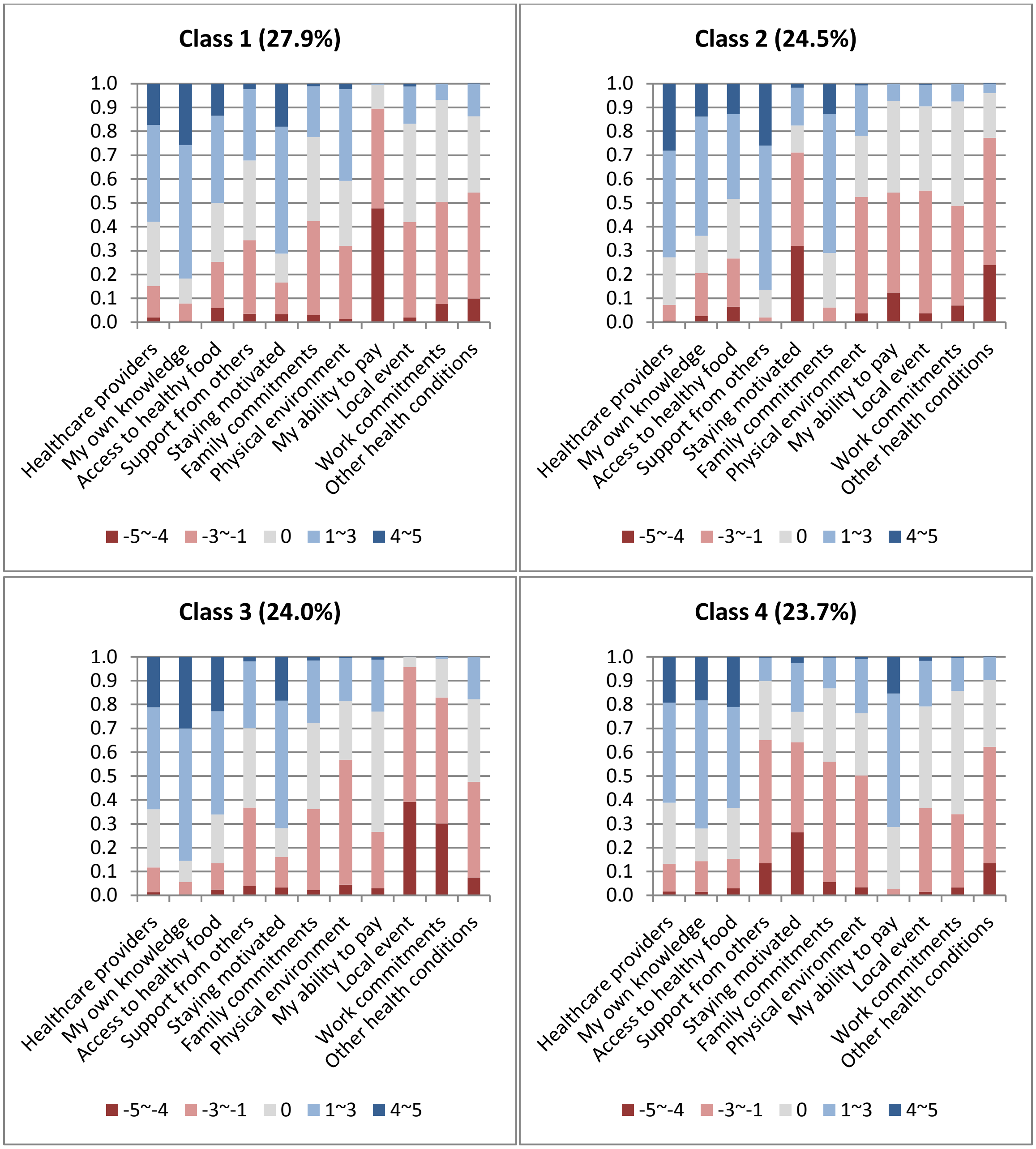


Table 4-1 Prioritization factors included in the best worst scaling survey

Local events	Do your local events (e.g. cultural, community, or religious) impact your ability for diabetes self-management?
Support from others	Do you have enough support from friends, co-workers, support groups or others in your community?
Access to healthy food	Do you have regular access to healthy food that will support your ability for diabetes self-management?
Healthcare providers	Do your healthcare providers have a positive or negative impact on your ability to self-manage you diabetes?
Physical environment	Does the place/location where you live and work provide you with the resources to manage your diabetes?
Staying motivated	Do you usually have the self-control to make the best choices for managing your diabetes?
My own knowledge	Do you feel you know enough about diabetes to self-manage your diabetes?
Family commitments	Does your family have a positive or negative impact on your ability to self-manage your diabetes?
Work commitments	Does your work (or other responsibilities) affect your ability to self-manage your diabetes?
My ability to pay	Do you have enough money to successfully self-manage your diabetes?
Other health conditions	Do you have other health conditions (mental and physical) that affect how you manage your diabetes?

Table 4-2 Sample characteristics

	N	Percent
Demographic Characteristics		
Age (mean)	536	61.66
Gender		
Male	285	0.53
Female	251	0.47
Race		
White	280	0.52
Black	121	0.23
Hispanic	116	0.22
Other	19	0.03
Education		
Less than high school	35	0.07
High school	335	0.62
Bachelor's degree or higher	166	0.31
Income		
< 25,000	133	.25
25,000 – 50,000	150	.28
50,000 – 74,999	110	.20
≥ 75,000	143	.27
Health-related Characteristics		
Years of diagnosis (mean)	530	11.17
Self-Reported Health		
Excellent	25	.05
Very good	141	.26
Good	230	.43
Fair	116	.22
Poor	24	.04
No. A1c measured in last 6 months		
None	41	.08
1 time	244	.45
> 2 times	239	.45
Don't know	11	.02
Most recent A1c level		
≥ 7.0%	230	.43
< 7.0%	212	.40
Don't know	89	.17
Type of diabetes medicine used		
No prescription medicine	44	.08
Only pills	325	.61
Only injections/shots	39	.07
Pills and injections/shots	127	.24
No complications	315	.59
No other chronic conditions	80	.15
Personality Characteristics		
Agree or strongly agree with the following statement		
I'm always optimistic about my future	334	.62
I have a lot of self-control	270	.50
I'm actively working to improve my health	333	.62
I consider myself a risk-taker	89	.17

Table 4-3 Multinomial logit regression of facilitator class membership on individual characteristics

	Marginal Effect (SE)				
	Class 1 Provider-driven	Class 2 Support-driven	Class 3 Healthy food-driven	Class 4 Motivation-driven	Class 5 Affordability-driven
Age	**0.005**** (.002)	-0.001 (.002)	-0.002 (.002)	-0.001 (.002)	-0.001 (.002)
Gender					
Male	*Reference*				
Female	-0.069 (.046)	0.042 (.044)	0.059 (.045)	0.034 (.038)	-0.065 (.035)
Race					
White	*Reference*				
Black	-0.080 (.054)	-0.058 (.046)	**-0.091*** (.046)	0.102 (.059)	**0.127*** (.056)
Hispanic	**-0.170**** (.056)	0.094 (.076)	0.014 (.060)	0.082 (.059)	-0.019 (.055)
Other	0.077 (.119)	-0.033 (.086)	-0.132 (.081)	-0.088 (.054)	0.178 (.107)
Education					
No high school	*Reference*				
High school	0.019 (.096)	**0.130*** (.057)	0.077 (.059)	-0.159 (.092)	-0.066 (.093)
Bachelor+	0.014 (.104)	**0.151*** (.072)	0.074 (.073)	-0.094 (.100)	-0.145 (.095)
Income					
< 25,000	*Reference*				
25,000-49,999	-0.047 (.063)	-0.054 (.065)	0.075 (.055)	0.030 (.053)	-0.003 (.057)
50,000-74,999	0.022 (.070)	-0.096 (.065)	0.085 (.064)	0.059 (.063)	-0.069 (.046)
≥ 75,000	-0.047 (.066)	-0.099 (.068)	0.064 (.064)	0.024 (.055)	0.057 (.059)
Health Status					
Excellent	*Reference*				
Very good	-0.070 (.105)	**0.120*** (.058)	-0.093 (.105)	0.103 (.079)	-0.059 (.100)
Good	-0.088 (.104)	**0.180**** (.054)	-0.038 (.102)	0.025 (.076)	-0.078 (.098)
Fair	0.011 (.114)	**0.173*** (.069)	-0.107 (.108)	-0.006 (.086)	-0.071 (.102)
Poor	**-0.240*** (.122)	**0.180*** (.096)	-0.004 (.138)	0.144 (.125)	-0.081 (.121)
No. HbA1c Test in past 6 months					
None	*Reference*				
1 or more	0.129 (.071)	0.096 (.058)	-0.147 (.089)	0.013 (.081)	-0.091 (.072)
Don't know	0.122 (.149)	-0.025 (.104)	0.252 (.164)	-0.135 (.086)	**-0.213**** (.067)
Recent HbA1c Level					
> 7%	*Reference*				
≤ 7%	-0.025 (.049)	0.019 (.044)	0.065 (.046)	-0.042 (.040)	-0.017 (.033)
Don't know	-0.086 (.063)	0.022 (.066)	0.041 (.061)	-0.012 (.069)	0.035 (.052)
I have a lot of self-control					
Disagree	*Reference*				
Agree	-0.078 (.060)	-0.010 (.052)	0.017 (.050)	**0.118*** (.052)	-0.047 (.050)
I consider myself a risk-taker					
Disagree	*Reference*				
Agree	**-0.174**** (.053)	**0.155*** (.072)	-0.028 (.054)	0.009 (.054)	0.038 (.058)

Table 4-4 Multinomial logit regression of barrier class membership on individual characteristics

	Marginal Effect (SE)				
	Class 1 Work-committed	Class 2 Affordability-lacking	Class 3 Motivation-lacking	Class 4 Support-lacking	Class 5 Healthy food-lacking
Age	-0.002 (.003)	0.001 (.002)	0.002 (.002)	-0.002 (.002)	0.000 (.002)
Gender					
Male	*Reference*				
Female	-0.014 (.051)	-0.013 (.043)	0.042 (.040)	-0.007 (.035)	-0.008 (.028)
Race					
White	*Reference*				
Black	0.039 (.065)	-0.011 (.057)	-0.059 (.055)	0.000 (.042)	0.032 (.040)
Hispanic	0.138 (.086)	0.020 (.062)	-0.085 (.065)	-0.006 (.050)	**-0.066**** (.023)
Other	0.035 (.106)	**-0.193***** (.042)	-0.039 (.078)	0.061 (.093)	0.137 (.073)
Education					
No high school	*Reference*				
High school	0.035 (.107)	-0.118 (.099)	**0.133*** (.065)	-0.074 (.085)	0.023 (.048)
Bachelor+	0.029 (.115)	-0.075 (.112)	**0.169*** (.075)	-0.095 (.096)	-0.028 (.050)
Income					
< 25,000	*Reference*				
25,000-49,999	-0.053 (.072)	0.017 (.065)	**0.117*** (.055)	-0.025 (.054)	-0.056 (.036)
50,000-74,999	0.053 (.079)	-0.074 (.067)	**0.131*** (.060)	-0.043 (.056)	-0.067 (.040)
≥ 75,000	-0.072 (.071)	-0.126 (.067)	**0.119*** (.062)	0.044 (.067)	0.034 (.045)
Employment					
Employed	*Reference*				
Unemployed	-0.042 (.070)	0.050 (.065)	-0.082 (.057)	0.037 (.054)	0.036 (.037)
Retired	-0.023 (.069)	-0.066 (.054)	0.030 (.061)	0.029 (.056)	0.030 (.043)
Health Status					
Excellent	*Reference*				
Very Good	0.097 (.108)	-0.007 (.095)	0.036 (.083)	0.123 (.064)	**-0.248*** (.111)
Good	0.028 (.105)	0.027 (.092)	0.076 (.084)	0.087 (.059)	**-0.217*** (.106)
Fair	-0.033 (.114)	0.029 (.098)	**0.239*** (.098)	0.003 (.061)	**-0.237*** (.112)
Poor	-0.162 (.131)	0.091 (.133)	0.261 (.155)	0.054 (.089)	**-0.244*** (.123)
No. HbA1c tests in past 6 months					
None	*Reference*				
1 or more	0.024 (.096)	-0.028 (.084)	-0.047 (.079)	0.051 (.050)	0.000 (.045)
Don't know	-0.020 (.162)	0.112 (.149)	-0.054 (.145)	0.054 (.123)	**-0.092*** (.040)
Recent HbA1c Level					
> 7.0%	*Reference*				
≤ 7.0%	0.041 (.054)	-0.045 (.046)	0.034 (.045)	0.026 (.040)	**-0.056*** (.024)
Don't know	0.007 (.074)	-0.087 (.053)	-0.014 (.064)	-0.030 (.044)	**0.124*** (.059)
I have a lot of self control					
Disagree	*Reference*				
Agree	**0.160*** (.067)	-0.052 (.061)	-0.111 (.061)	0.023 (.049)	-0.020 (.037)
I consider myself a risk-taker					
Disagree	*Reference*				
Agree	0.037 (.071)	0.085 (.066)	**-0.141**** (.051)	-0.002 (.056)	0.020 (.040)

Table 4-5 Multinomial logit regression of class membership from best-worst scores on individual characteristics

	Marginal Effect (SE)			
	Class 1	**Class 2**	**Class 3**	**Class 4**
Age	-0.001 (.003)	-0.001 (.002)	0.004 (.002)	-0.003 (.003)
Gender				
Male	*Reference*			
Female	-0.001 (.047)	-0.036 (.042)	-0.001 (.046)	0.038 (.047)
Race				
White	*Reference*			
Black	-0.006 (.060)	-0.013 (.058)	0.000 (.055)	0.019 (.069)
Hispanic	0.028 (.068)	0.050 (.068)	0.022 (.069)	-0.100 (.066)
Other	-0.072 (.092)	0.082 (.106)	0.006 (.098)	-0.016 (.114)
Marital Status				
Married	*Reference*			
Widowed/Divorced /Separated	0.056 (.059)	**-0.138**** (.049)	0.025 (.052)	0.056 (.060)
Never married	0.117 (.081)	**-0.227***** (.048)	0.081 (.078)	0.029 (.075)
Education				
No high school	*Reference*			
High school	**-0.241*** (.116)	**0.234***** (.042)	-0.025 (.097)	0.033 (.110)
Bachelor+	**-0.274*** (.129)	**0.219***** (.057)	0.086 (.106)	-0.032 (.121)
Income				
< 25,000	*Reference*			
25,000-49,999	0.009 (.068)	0.033 (.069)	-0.051 (.061)	0.009 (.067)
50,000-74,999	-0.056 (.075)	0.006 (.070)	0.085 (.073)	-0.035 (.067)
≥ 75,000	-0.098 (.079)	-0.048 (.073)	-0.015 (.070)	**0.162*** (.079)
Employment				
Employed	*Reference*			
Unemployed	0.058 (.068)	-0.036 (.060)	0.007 (.067)	-0.029 (.061)
Retired	-0.043 (.062)	0.019 (.065)	-0.038 (.059)	0.062 (.071)
Health Status				
Excellent	*Reference*			
Very Good	-0.026 (.110)	0.058 (.073)	-0.021 (.097)	-0.011 (.119)
Good	-0.057 (.108)	**0.197**** (.072)	-0.073 (.097)	-0.067 (.119)
Fair	-0.057 (.116)	**0.187*** (.086)	-0.108 (.107)	-0.022 (.132)
Poor	0.020 (.147)	**0.306*** (.130)	-0.189 (.130)	-0.138 (.144)
No. HbA1c tests in past 6 months				
None	*Reference*			
1 or more	-0.007 (.083)	0.072 (.077)	0.011 (.083)	-0.076 (.088)
Don't know	0.159 (.162)	-0.119 (.099)	0.164 (.174)	-0.204 (.130)
Recent HbA1c Level				
> 7.0%	*Reference*			
≤ 7.0%	**-0.104*** (.048)	0.073 (.043)	0.058 (.049)	-0.028 (.048)
Don't know	-0.107 (.069)	0.129 (.076)	0.010 (.060)	-0.032 (.064)
I have a lot of self-control				
Disagree	*Reference*			
Agree	0.002 (.062)	-0.031 (.057)	**0.167**** (.053)	**-0.139*** (.066)

Table 4-6 Cross-tab between class membership from best-worst scores and class membership from best or worst choices

		BW Score				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Total**
Facilitator Classes	**Class 1**	37	39	33	33	142
	Provider-driven	26.1%	27.5%	23.2%	23.2%	
	Class 2	25	74	15	2	116
	Support-driven	21.6%	63.8%	12.9%	1.7%	
	Class 3	29	18	36	31	114
	Health food-driven	25.4%	15.8%	31.6%	27.2%	
	Class 4	55	0	40	10	105
	Motivation-driven	52.4%	0	38.1%	9.5%	
	Class 5	6	7	7	39	59
	Affordability-driven	10.2%	11.9%	11.9%	66.1%	
	Total	152	138	131	115	536
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Total**
Barrier Classes	**Class 1**	22	30	108	14	174
	Work-committed	12.6%	17.2%	62.1%	8.0%	
	Class 2	94	28	9	0	131
	Affordability-lacking	71.8%	21.4%	6.9%	0	
	Class 3	4	55	2	53	114
	Motivation-lacking	3.5%	48.2%	1.8%	46.5%	
	Class 4	17	2	10	43	72
	Support-lacking	23.6%	2.8%	13.9%	59.7%	
	Class 5	15	23	2	5	45
	Healthy food-lacking	33.3%	51.1%	4.4%	11.1%	
	Total	152	138	131	115	536

Chapter 5 Conclusion and Policy Implication

Stated-preference researchers have been increasingly using latent class analysis (LCA) to analyze preference heterogeneity in the health context in the past decade in a widening range of countries. Preferences have been assessed on a variety of topics, including medical treatments, preventive interventions, general health services, quality of life, provider job preference, and resource allocation. Patient preference is the focus in majority of the research. Several studies compared how preference varied from different perspectives. While most segmentation was performed to direct the development of tailored treatments, prevention programs, or personalized care in the past twenty years, results on preference heterogeneity have been increasingly used to inform policy since 2010.

Although the quality of LCA studies has been improving over time, there is still large variation in how the analysis is conducted. Many studies failed to justify the choice of sample size, model specification, model selection criteria, and individual covariates to include in the model, which seem to be largely determined by the software used for LCA. However, an analysis across the studies showed that the number of classes reported in the study was not correlated with stated-preference methods, segmentation methods, or model selection criteria, but only sample size. This could be because the number of classes identified in a study was indeed not subject to survey or segmentation methods, but given that 80% of the studies reported two to three classes, it is likely that most studies were underpowered to detect more classes. In addition, many studies limited the number of classes they report by considering the interpretation of the segments when selecting models, especially when information criteria or other statistical criteria failed to identify a practical amount of segments for implementation. Although policy and clinical relevance is important, model fit should still be considered to ensure reliable estimates and accurate prediction.

Latent class logit (LCL) is the most commonly used segmentation model for discrete choice data. It has been applied in preference data obtained from discrete choice experiment (DCE) and best-worst scaling (BWS), two most common stated-preference methods used in health according to the systematic review. However, both applications have limitations. In DCE, when there is substantial within-class preference heterogeneity or significant overlap between classes, LCL may over-fit the data and lead to too many classes that are difficult to incorporate in policy or clinical decision-making. As a remedy, random effects can be incorporated in LCL models. With more flexible model specification, random effect LCL (RELCL) is shown not only reducing the number of classes but also better capturing the complex and dispersed preference pattern among patients than LCL, leading to improved model fit and prediction accuracy. Given that healthcare regulatory and clinical decision-makers often prefer parsimonious results due to limited resources and capacity to accommodate too many preference types, researchers can consider using RELCL as a tool to generate more practical results to meet the specific needs.

When LCL is applied to BWS data, information criteria also often fail to identify the best-fitted model and parsimonious segmentation results. This is because the utility parameter for each object in LCL is assumed to be constant across choices tasks and therefore sensitive to the full sequence of choices. When there are many choice tasks and large variation in choice sequence among respondents, there is significant variation in utility parameters. Segmentation based on a large set of utility parameters with substantial variances can lead to a large number of classes. By relaxing the cross-task constant utility constraint, the standard LC model can dramatically reduce the number of classes and generate parsimonious segmentation results. Depending on the study objective, the standard LC model can either be applied to the best or worst choices, or to the best-worst scores. Overall, researchers should consider using flexible model specification for reliable and practical results to support policy and clinical decision-making when analyzing BWS data.

5.1 Policy and Clinical Implications

In 2012, the President signed into law the Food and Drug Administration Safety and Innovation Act (FDASIA) of 2012, which reauthorized the Prescription Drug User Fee Act (PDUFA). Subsequent PDUFA amendments require the Food and Drug Administration (FDA) to develop and implement a structured approach to benefit-risk assessment in the new drug-approval process. In response to the amendments, the FDA's Center for Drug Evaluation and Research (CEDR) established the Patient-Focused Drug Development initiative[2], aiming to obtain patients' perspective on disease conditions and treatments using a systematic approach. FDA launched the initiative in 2012 by announcing 20 disease areas to focus on during the first three years[3]. Public meetings were conducted among FDA review divisions, patient-advocacy communities, and other interested stakeholders to gather information on the challenges and possible barriers to treatment patients face in the selected disease areas to help inform drug-development and regulatory review processes.

---

[2] Patient-Focused Drug Development. http://www.fda.gov/ForIndustry/UserFees/PrescriptionDrugUserFee/ucm326192.htm.

[3] Prescription Drug User Fee Act Patient-Focused Drug Development; Announcement of Disease Areas for Meetings Conducted in Fiscal Years 2016–2017. http://www.gpo.gov/fdsys/pkg/FR-2015-07-02/pdf/2015-16359.pdf

During the same year, the FDA Center for Devices and Radiological Health (CDRH) released guidance to clarify the principal benefit-risk factors that FDA considers during the reviews for premarket approval for medical devices (CDRH, 2012). In addition to description of benefits and risks, the guidance listed patient preference as an important factor to consider in regulatory reviews. It recognizes heterogeneity in risk tolerance and benefit-risk tradeoff among patients, and expresses the willingness to consider the whole spectrum of patient preferences and approve a medical device that demonstrates meaningful benefits even though its benefit-risk profile is only acceptable to a subset of risk-tolerant patients. A more recent guidance provides details on how patient preference information should be included in the applications (CDRH, 2015).

Learning patient preference and preference heterogeneity will support these policy movements towards patient-centered regulatory decision-making and expand access to effective treatments for patients. Patients have unique perspectives about the benefits and risks of medical treatments. But in order to incorporate these views in regulatory decision-making, researchers should have reliable and accurate methods, tools, and approaches (Johnson et al, 2016). Stated-preference methods have been widely used in health to quantify patients' opinions. Compared to qualitative approaches such as focus group interviews, they allow researchers to survey a larger sample of patients and study heterogeneity in preferences. As segmentation is an appealing approach to describe preference heterogeneity due to the straightforward interpretation and policy implication, its application in stated-preference studies in health is often unsatisfactory. We proposed alternative models in this dissertation to improve such application. Future researchers can consider using these models to produce more reliable, accurate, and practical results to support regulatory decision-making.

The diabetes case study demonstrated how these models could be used to support patient-centered drug development. Significant preference heterogeneity for diabetes medication was found among patients. While some patients focused on cost and effectiveness of the medicine when choosing treatment, others cared more about side effects and treatment burden. Such heterogeneity was not only associated with patients' socio demographic characteristics, such as age and income, but also affected by their health status, current treatment, and even personality. To meet the various needs, FDA may consider approving more effective or cheaper medicines for people who are willing to take higher risk of side effects or greater treatment burden, and approving medicines with fewer side effects for those who want to avoid side effects even for a higher price. Patients should have access to both to choose from according to their preference.

In addition to preference heterogeneity in treatment, barriers and facilitators for diabetes management also varied among patients, which can be explained by socio demographics, health status, and personality. Although healthcare providers and policy makers have been emphasizing expanding access to healthy food and holding local events to improve patients' diabetes self-management, we found that access to healthcare, family and social support, work commitment and motivation had more significant impact on diabetes management. Expanding health insurance coverage to the uninsured or underinsured patients and enhancing work environment to facilitate disease management are more effective approaches that policy makers can take to improve health outcomes among diabetes patients. Healthcare providers can consider involving patients' family and other social groups during the treatment to support and motivate patients. A collaborative effort between regulators and providers to develop treatment plans that are tailored to meet individual needs is likely to improve treatment adherence and health outcomes.

5.2 Limitations

There are several limitations in the study. First, the datasets used in this dissertation are relatively small. There has not been recommendation on sample size calculation for LCA, because the calculation depends on the number of classes, difference in utility parameters between classes, and the class size, all of which is difficult to predict. Orme (2006) recommended a rule of thumb of 200 respondents per subgroup, but since LCA is probabilistic (i.e., each respondent has a probability of belonging to each class), 200 per subgroup may not be necessary. In addition, as the classes may not have equal share of the study sample, it is possible that some groups end up having fewer than 200 respondents. With 500 respondents, although our sample size is larger than majority of the segmentation studies in the literature, it still limited our ability to impose more flexible model specifications with more parameters. Future studies should use larger dataset to test the models we proposed in this dissertation.

Second, the sample was not representative of the entire diabetes population. As discussed in the previous chapters, the study samples were more educated and had higher income on average than the general diabetes patient population. Although sampling weight was used to adjust standard errors and class distribution, the lower-income and less-educated patients' preference and priority was underrepresented in the sample. This is particularly important when we study preference heterogeneity. Most studies in the literature do not examine the representativeness of the sample. Researchers and decision-makers therefore should be cautious about generalizing results to the entire study population or comparing across studies. Representative study sample should be emphasized for future stated-preference studies in health to better support policy.

Third, the proposed models were only applied to preference data from patients with type 2 diabetes. Since diabetes is a prevalent chronic condition, diabetes patient group can be more diverse than other patient groups in terms of socio demographics, health status, experience, culture, and personalities. The diverse population characteristics can lead to more dispersed preference distribution that we may not see among patients with more acute or severe conditions. The proposed models can be tested among other patients groups in future research. A comparison of preference distribution between different patient groups can be beneficial.

Fourth, we only used patients' stated-preference data, the choices in which are likely to be different from the choices observed in reality. However, because stated-preference methods measure patient preference in controlled laboratory environment and remove confounding from other factors, they better serve FDA's goal of understanding patient perspectives on benefits and risks of medical treatments. Researchers could test the proposed models in revealed preference data, but the models should be adjusted to control for confounding from such factors as provider preference, choice context, and choice set.

Finally, one weakness of testing models using empirical data is that the true preference parameters are unknown. We compared alternative models with different distribution assumptions and assessed their fit using information criteria, but we were unable to examine how far the estimates from the best-fitted model were from patients' true utility values. Researchers should use simulation to further evaluate the performance of the proposed models in this dissertation and assess the threshold for when the proposed models start to outperform the traditional ones.

Despite the limitations, this dissertation overviews the application of LCA in stated-preference studies in health. It identifies the issues in the application and modifies the traditional segmentation models to better fit the complex preference data in health. By allowing more flexible model specifications, we provide stated-preference researchers a tool to generate parsimonious and practical results to serve regulatory decision-making purposes. This dissertation can serve as a methodological guide for researchers who are interested in studying preference heterogeneity in health.

Reference

Califf RM. Benefit-risk assessments at the US Food and Drug Administration: finding the balance. JAMA. 2017. doi: 10.1001/jama.2017.0410. [Epub ahead of print]

Center for Devices and Radiological Health (CDRH) and Center for Biologics Evaluation and Research (CBER). Guidance for industry and Food and Drug Administration staff: factors to consider when making benefit-risk determinations in medical device premarket approval and de novo classifications. Issued on March 28, 2012.

Center for Devices and Radiological Health (CDRH) and Center for Biologics Evaluation and Research (CBER). Patient preference information – submission, review in PMAs, HDE applications, and *de novo* requests, and inclusion in device labeling: draft guidance for industry, Food and Drug Administration staff, and other stakeholders. Issued on May 18, 2015.

Food and Drug Administration Safety and Innovation Act of 2012, 21 U.S.C. § 3187.

Johnson R, and Zhou M. Patient preferences in regulatory benefit-risk assessments: a U.S. perspective. Value in Health. 2016;19(6):741-5

Appendix: CV

MO ZHOU

PERSONAL DATA

Birth Date: February 16, 1985
Birth City: Qinhuangdao, China
Home Address: 675 President St., Suite 1604, Baltimore, MD 21202
Phone: +1 (917) 361-3724
Email: mzhou16@gmail.com

Work Address:
Center for Health Services and Outcomes Research (CHSOR)
Department of Health Policy and Management
624 N. Broadway, Room 690
Johns Hopkins Bloomberg School of Public Health
Phone: +1 (443) 287-1779
Email: mzhou8@jhu.edu

EDUCATION AND TRAINING

Ph.D. in Health Economics and Policy (2013-2017)
Johns Hopkins Bloomberg School of Public Health
Dissertation title: Explore heterogeneity in stated preference through latent class analysis
Committee members: John F.P. Bridges, PhD; Darrel Gaskin, PhD; Karen Bandeen-Roche, PhD; and Jodi Segal, MD, MPH

Master of Health Science (MHS) in Health Economics (2012-2013)
Johns Hopkins Bloomberg School of Public Health

Predoctoral fellowship in Health Economics (2011-2012)
University of Minnesota School of Public Health

Master Public Administration (MPA) in Health Policy and Management (2008-2010)
Wagner School of Public Affairs, New York University

Bachelor of Medicine in Nursing/Bachelor of Economics (2003-2008)
Peking University

HONORS AND AWARDS

Teaching Assistant Recognition Award, Johns Hopkins Bloomberg School of Public Health, 2016
Dean's Scholarship, University of Minnesota School of Public Health, 2011
Medtronic Interdisciplinary Healthcare Case Competition, Finalist, 2011

PROFESSIONAL EXPERIENCE

Research Assistant (2012-current)
Center for Health Services and Outcomes Research (CHSOR)
Department of Health Policy and Management, Johns Hopkins School of Public Health

- Conduct systematic review of research in economics and medicine
- Perform econometric analysis on original and secondary data
- Conduct cost-effectiveness analyses
- Draft and revise manuscripts for publication
- Prepare power-point presentations and posters for seminars and conferences
- Organize seminars and local events with local community members

Research Intern (2012)
Clinton Health Access Initiative (CHAI)

- Estimated the impacts of human resource arrangement on HIV treatment costs and health outcomes in five African countries to draw policy recommendations to improve the efficiency of human resources allocation
- Proposed a discrete choice experiment to assist Ministry of Health in Zambia in identifying cost-effective retention strategies to attract health workers in underserved rural areas

Research Analyst (2009-2011)
Department of Environmental and Occupational Health
NYC Department of Health and Mental Hygiene

- Assisted climate epidemiologists in the development and implementation of the Climate Change Adaptation Program in NYC
- Supported an epidemiology study on drowning and the development of evidence-based drowning prevention and outreach policies in NYC
- Collaborated with epidemiologists in surveillance for food-borne illnesses in NYC

Project Associate (2009-2010)
Weill Cornell Medical Center of New York-Presbyterian Hospital

- Collaborated with hospital administrators to evaluate the impact of the automated medication distribution system on patient safety and health care quality
- Collected primary qualitative and quantitative data through interviews with nurses and pharmacists and observations of drug dispensing
- Developed a work plan for an optimal implementation of the automated medication distribution system in the hospital

Intern Nurse (2007-2008)
Peking University First Hospital

- Provided medical care to patients in both inpatient and outpatient settings
- Assisted in nursing research on health care quality improvement
- Supported the nurse managers in daily unit management

PROFESSIONAL ACTIVITIES

Society Membership
International Health Economics Association (IHEA)
International Society for Pharmacoeconomics and Outcome Research (ISPOR)
Society of Medical Decision-Making (SMDM)

PUBLICATIONS

Journal Articles
Johnson R, **Zhou M**. (2016). Patient preferences in regulatory benefit-risk assessments: a U.S. perspective. *Value in Health*, 19(6): 741-5.

Xu T, Makary, Al Kazzi E, **Zhou M**, Paulik TM, Hutfless SM. (2016). Surgeon-level variation in postoperative complications. *Journal of Gastrointestinal Surgery*, 20(7): 1393-9.

Mogul D, **Zhou M**, Intihar P, Schwarz K, Frick K. (2015). Cost-effective analysis of screening for biliary atresia with the stool color card. *Journal of Pediatric Gastroenterology and Nutrition*, 60(1): 91-8.

Xu T, Hutfless SM, Cooper MA, **Zhou M**, Massie AB, Makary MA. (2015). Hospital cost implications of increased use of minimally invasive surgery. *JAMA Surgery*, 150(5): 489-90.

Book Chapters
Johnson R, **Zhou M**. (2016). Quantifying patient preferences for regulatory benefit-risk assessments. In Q. Jiang, W. He (Eds.), *Benefit-risk assessment methods in medical product development: bridging qualitative and quantitative assessments*. Taylor & Francis Group.

Working Papers
Zhou M, Thayer W, Bridges JFP. Using latent class analysis to model preference heterogeneity in health: a systematic review.

Zhou M, Bridges JFP. Explore preference heterogeneity for the treatment of people with type 2 diabetes: a comparison of random-parameters and latent-class estimation techniques.

Zhou M, Bandeen-Roche K, Bridges JFP. Analyzing preference heterogeneity in best-worst scaling data: a comparison of latent class logit and standard latent class models.

Zhou M, Bandeen-Roche K, Bridges JFP. Using latent class analysis to analyze heterogeneity in barriers and facilitators for diabetes self-management among patients with type 2 diabetes.

Zhou M, Bridges JFP, Padula WV, Segal J. Exploring potential drivers of the systematic overuse of healthcare in the United States using the Johns Hopkins Overuse Index (JHOI).

Bridges JFP, **Zhou M**, Segal J, Bandeen-Roche K, Wu A. A comparison of Likert and best-worst scaling for measuring patient priorities: a randomized controlled trial.

TEACHING

Guest Lecturer

Mathematical Microeconomics, Johns Hopkins School of Public Health (2015-16, 2016-17)
Primary Instructor: John F.P. Bridges

Economic Evaluation III: Conjoint Analysis, Johns Hopkins School of Public Health (2015-16)
Primary Instructor: Gregory de Lissovoy

Teaching Assistant

Health Economics I, Johns Hopkins School of Public Health (2013-14, 2015-16)

Teaching Assistant to Douglas Hough

Health Economics II, Johns Hopkins School of Public Health (2015-16)
Teaching Assistant to Matthew Eisenberg

Health Economics III, Johns Hopkins School of Public Health (2014-16)
Teaching Assistant to David Bishai

Economic Evaluation II: Markov Models, Johns Hopkins School of Public Health (2014-17)
Teaching Assistant to William Padula

Economic Evaluation III: Conjoint Analysis, Johns Hopkins School of Public Health (2013-14)
Teaching Assistant to John F.P. Bridges

RESEARCH GRANT PARTICIPATION

Advancing Stated-Preference Methods for Measuring the Preferences of Patients with Type 2 Diabetes (2014-2016)
Sponsoring agency: Patient-centered outcome research institute (PCORI)
PI: John F.P. Bridges
Role: Research Assistant

- Study patient priority in barriers and facilitators for diabetes self-management using segmentation methods
- Study patient preference and preference heterogeneity over diabetes medications using discrete choice models

ACADEMIC SERVICE

Student Coordination Committee (2014-2015), Department of Health Policy and Management
Johns Hopkins Bloomberg School of Public Health

PRESENTATIONS

Scientific Meetings

Analyzing preference heterogeneity in best-worst scaling data: a comparison of latent class logit and standard latent class models

- International Health Economics Association (iHEA) Congress [Boston], July 2017 (podium)

Using latent class analysis to model preference heterogeneity in health: a systematic review.

- International Society for Pharmacoeconomics and Outcome Research (ISPOR) Annual International Meeting [Boston], May 2017 (poster)

Using latent class analysis to analyze heterogeneity in barriers and facilitators for diabetes self-management among patients with type 2 diabetes.

- International Society for Pharmacoeconomics and Outcome Research (ISPOR) Annual International Meeting [Boston], May 2017 (poster)

Explore preference heterogeneity for the treatment of people with type 2 diabetes: a comparison of random-parameters and latent-class estimation techniques

- International Choice Modeling Conference (ICMC) 2017 [Cape Town, South Africa], April 2017 (podium)

A comparison of Likert and best-worst scaling for measuring patient priorities: a randomized controlled trial

- Society of Medical Decision-Making (SMDM) 38th Annual Meeting [Vancouver], October 2016 (podium)

Exploring potential drivers of the systematic overuse of healthcare in the United States using the Johns Hopkins Overuse Index (JHOI)

- International Society for Pharmacoeconomics and Outcome Research (ISPOR) Annual International Meeting [Philadelphia], May 2015 (poster)
- International Health Economics Association (iHEA) Congress [Milan], July 2015 (podium)

Seminars

Exploring heterogeneity in stated preference among patients through latent class analysis

- Center for Health Services and Outcomes Research (CHSOR), Johns Hopkins Bloomberg School of Public Health, October 2016

ADDITIONAL INFORMATION

Research objectives

Mo Zhou is currently a Ph.D. candidate in the Health Economics and Policy program at the Johns Hopkins Bloomberg School of Public Health. With extensive training in econometrics, her research focuses on using discrete choice analysis to estimate demand for health care and the application of conjoint analysis in medicine. She has been involved in stated-preference studies in areas such as diabetes and vaccination and is particularly interested in studying preference heterogeneity and market segmentation using various econometric techniques.

Keywords

Choice modeling, econometrics, stated-preference methods, conjoint analysis, patient preference, demand for health care, economic evaluation